



HAL
open science

Expert judgments calibration and combination for assessment of river levee failure probability

Michel Hathout, Marc Vuillet, C. Carvajal, L. Peyras, Youssef Diab

► To cite this version:

Michel Hathout, Marc Vuillet, C. Carvajal, L. Peyras, Youssef Diab. Expert judgments calibration and combination for assessment of river levee failure probability. *Reliability Engineering and System Safety*, 2019, 188, pp.377-392. 10.1016/j.ress.2019.03.019 . hal-02609390

HAL Id: hal-02609390

<https://hal.inrae.fr/hal-02609390>

Submitted on 22 Oct 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

Expert judgments calibration and combination for assessment of river levee failure probability

Michel HATHOUT*/**, Marc VUILLET*, Claudio CARVAJAL**, Laurent PEYRAS**, Youssef DIAB*/***.

* Lab'Urba, Ecole des ingénieurs de la ville de Paris, Université Paris-Est, 80 rue Rébeval, 75019 Paris - France

michel.hathout@eivp-paris.fr ; marc.vuillet@eivp-paris.fr

** Irstea, unité de recherche RECOVER 3275 route Cézanne – CS 40061 – 13182 Aix en Provence - France

claudio.carvajal@irstea.fr ; laurent.peyras@irstea.fr ; michel.hathout@irstea.fr

*** Lab'Urba, Université Paris-Est (Marne-la-Vallée), bureau J29, Bât Lavoisier, 5 Bvd Descartes 77420 Champs-sur-Marne - France

youssef.diab@u-pem.fr

Abstract

In civil engineering, the utilisation of expert judgement is important for evaluating the reliability of structures whose failure behaviour is little known and difficult to quantify. This article presents the application of the approaches of elicitation, calibration and aggregation of expert opinions to evaluate the reliability of river levees. These approaches have seldom been used in the field of French hydraulic structures whereas there is growing interest for them in other countries, concerning structures reliability and other fields such as aerospace industry, nuclear industry, hydrology, statistics, economics, psychology etc. This article proposes a quantitative approach to the use of expert judgement for evaluating river levee failure probability. An application to the case of an existing river levee is presented. The approach developed assesses the expert opinions according to their calibration (statistical accuracy) and their informativeness, then aggregates them in a single calibrated opinion. Applying the proposed weighting and aggregation procedure reduces variability in the probability of failure estimate, then it provides a better probabilistic distribution of aggregated expert opinion than of individual expert opinions. The results allow identifying a trend of aggregated expert opinions that point towards over- or lack of confidence.

Keywords: *structural reliability, subjective probability, levees, expert judgement, elicitation, calibration*

1 Introduction

River levees intended for flood protection are long linear structures of critical importance for people and property. Obtaining better knowledge of their risks and their structural reliability is a major challenge for engineers and the managers of these structures. However, the levees

37 in France and abroad are often old, little known, and poorly maintained and documented [1,
38 2]. Evaluating their risks and reliability is complex due to the small quantities of data
39 available in relation to the size and heterogeneity of these structures, and the complexity of
40 failure mechanisms. Indeed, there appears to be a general lack of statistical data on levee
41 characteristics (geometric, geotechnical and hydraulic data).

42 Furthermore some failure mechanisms such as internal erosion, overtopping and scouring are
43 subject to many stability criteria that can be found in the literature according to the type of
44 erosion, materials and levee profiles [2, 3, 4, 5, 6, and 7]. However it is difficult to make
45 formulation choices given the large number of physical laws and parameters potentially
46 combinable as well as uncertainties resulting from these modelling assumptions. There is not
47 really consensus on the limit state conditions for internal erosion, overtopping and scouring in
48 the national regulations, standards and professional recommendations. Only the sliding
49 mechanism is subject to a well-established limit state condition in the regulations and
50 recommendations [8].

51 In this context, evaluation using different quantitative approaches for risks and the structural
52 reliability of river levees regarding different failure mechanisms – sliding mechanism, internal
53 erosion, overtopping and scouring – is complex and often unfeasible [9]. This leads to calling
54 on expert judgement in risk analysis studies. Expert engineers must interpret the available
55 data and voice their uncertainties, in order to then evaluate the reliability of levees in terms of
56 failure probability. Moreover, during the assessment of the safety of the French levees, it is
57 necessary to mobilize various areas of expertise and complementary: hydraulic, morpho
58 dynamic, geotechnical, expertise of hydraulic structures, civil engineering, etc. Thus it is
59 necessary to call upon several experts, in charge of deciding on a probability of failure, in a
60 context of high uncertainties: about the physics of the phenomena considered, the
61 representativeness and the variability of the available data.

62 The ILH (International Levees Handbook) [2] underlines the advantage and importance of
63 using expert judgement for analysing the risks and reliability of levees when evaluation by
64 quantitative approaches is difficult to perform. However, it does not recommend any specific
65 approach allowing the efficient utilisation of expert judgement in an uncertain context. In the
66 scientific literature, there are few works focused on the use of expert judgement for analysing
67 the risks of levee failures. Perlea and Ketchum [10] proposed the use of expert judgment to
68 build fragility curves, assessing impact of vegetation, rodents' activity, encroachments, utility
69 penetrations, and erosion of the river bank and levee slopes. Serre *et al.* [11] proposed a
70 deterministic model of evaluating the performance of levees based on a multi-criteria decision
71 aid method: the rule-based assignment model. Vuillet *et al.* [12, 13] proposed a levee
72 evaluation model based on performance indicators, using the judgement of an expert or a
73 consensus of experts and including uncertainties on the data available on the levees. This
74 model is semi-quantitative and does not allow obtaining the failure probability of structures or
75 combining several expert opinions. It appears important to employ a methodology capable of
76 dealing with the latter two points, especially in France where the regulations relating to
77 structural hazard studies systematically require engineers to declare on the failure probability
78 of important structures.

79 Numerous approaches have been developed proposing probabilistic modelling of expert
80 assessments to study the reliability of structures [14, 15], in various fields such as the
81 aerospace industry, the nuclear industry, resistance of buildings to earthquakes [16] and
82 hydraulic structures [17]. Cooke [18] and Cooke and Goossens [19] distinguished three
83 approaches for the implementation of expert judgement: i) elicitation, ii) calibration and iii)
84 the aggregation of expert opinions.

85 The elicitation of expert opinion is a process of collecting expert judgements using specialised
86 verbal and written communication methods that allow quantifying risk [20]. There is general
87 scientific consensus regarding the main steps to be followed for the elicitation of expert
88 opinions, shared by several researchers in the domain like Morgan and Henrion [21]; Cooke
89 [18]; Clemen and Reilly [22], Ayyub [23]; Cooke and Goossens [19], Garthwaite *et al.* [24]
90 O'Hagan *et al* [14] and Burgman [25]. These steps generally correspond to the definition of
91 the problem, the identification and recruitment of experts and then the elicitation of expert
92 judgement.

93 To help the elicitation of experts in a probabilistic framework, economics researchers have
94 revealed the importance of the expert having access to data describing the context studied and
95 on which a probabilistic estimation can be made [26]. These data can be theoretical or
96 experimental. Furthermore, the authors show the importance of implementing the precise
97 modalities of eliciting expert opinions in a probabilistic format and recommend distinguishing
98 a range of values during elicitation and specifying that which seems most likely [27]. Relying
99 on O'Hagan *et al* [14], Mason *et al* [28] says that with individual elicitation, each expert gives
100 their view independently, and these views are expressed numerically and then appropriately
101 aggregated. The alternative consisting in an elicitation group in the form of an expert panel
102 allows experts to debate and reach consensus about the elicited values, but is more time
103 consuming and limits the number and range of experts included.

104 The calibration of expert opinions permits evaluating the consistency between the information
105 supplied by the expert and the known values (experimental / observed) [18]. The approach to
106 calibrating the expert opinions consists in evaluating and weighting the expert opinions in
107 relation to calibration variables whose real values are known. The approach entails
108 performing a quantitative evaluation of the calibration, or statistical accuracy, which refers to
109 how often experts capture the realization into their uncertainty interval and informativeness of
110 the expert elicitation through the use of calibration variables whose true values are known by
111 the analyst. The mathematical calibration models such as the classical model of Cooke [18]
112 allow determining a performance weight for expert proportional to the product of statistical
113 accuracy and informativeness. The classical model has been used in several applications [15,
114 29, 30], such as nuclear, banking, volcanoes, underground gas pipelines reliability [31, 32, 33,
115 34], to quantify internal erosion process of dams [35] and levees failure frequency in
116 Netherland [36]. One of the objectives of the calibration models is to aggregate expert
117 opinions onto a single opinion, however several authors recommend combining and
118 aggregating expert opinions using a simple average of the expert elicited values, which does
119 not require calibration [37, 38]. In a probabilistic framework, Lichtendahl *et al.* [39] proposed
120 a method for aggregating expert opinions based on the mean of quantiles elicited by the

121 experts, permitting aggregation in the form of an uncertainty interval and a most probable
122 value. Colson and Cooke [40] also proposed aggregating expert elicited distributions
123 constructed using the elicited values values by performance weighting.

124 The objective of this article is to present an approach for the elicitation, calibration and
125 aggregation of expert opinions for evaluating the probabilities of river levee failures due to
126 failure mechanisms. We wish to propose a method: i) to unify the estimates of several experts
127 and reduce the variability of the expert estimates, ii) to assess the global uncertainty of the
128 expert estimates and the degree of confidence of each expert, iii) and to assess the
129 uncertainties that each expert estimates and finally the degrees of confidence that can be given
130 to the expert estimates given their calibration scores.

131 The article first presents the development of an approach for eliciting expert opinions in terms
132 of failure probability represented by the slope sliding failure mechanism. It then presents the
133 application of Cooke's [18] calibration model to levees in order to weight the experts'
134 probabilistic opinions. Lastly, it presents the result of applying the protocol for aggregating
135 expert opinions developed by Lichtendahl *et al.* [39] to determine, on the basis of expert
136 judgement, a calibrated and aggregated probabilistic distribution of failure probabilities for
137 several profiles of an existing levee, using an aggregated interval to take into account the
138 uncertainties expressed by the experts.

139

140 2 *Methodological approaches for implementing expert judgement proposed* 141 *for river levees*

142 The studies of risk analyses and evaluations of structural reliability relating to levees must
143 include the evaluation of the probabilities of scenarios leading to failure [2]. The approaches
144 proposed focus on the elicitation of expert judgement, and the calibration and aggregation of
145 failure probabilities elicited relating to levee failure mechanisms that can lead to a breach of a
146 levee.

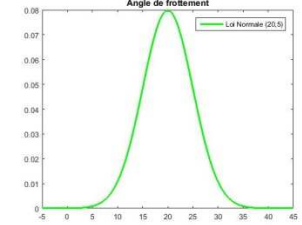
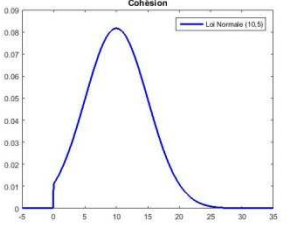
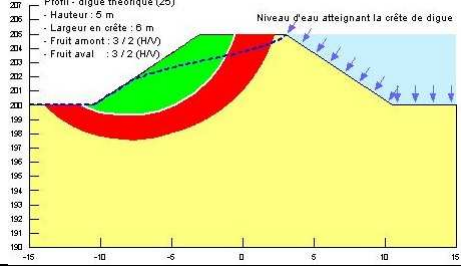
147 2.1 **Expert opinion elicitation approach**

148 The implementation of the expert opinion elicitation approach starts with the identification of
149 a panel of experts whose activities are directly involved with river levees. Levee risk analysis
150 studies involve a wide range of expert opinions linked to different disciplines: geotechnics,
151 river hydraulics, civil engineering, etc. Several experts from several disciplines with different
152 levels of experience (junior, confirmed, senior) are liable to intervene in these levee risk
153 analysis studies. Thus, we gathered an expert panel by taking into account the
154 interdisciplinary dimension of the different experts and their diversity of opinions [21].

155 Once the panel has been formed, each expert is questioned individually on the failure
156 probabilities relating to several profiles of the levee to be evaluated regarding different failure
157 mechanisms. A form including a questionnaire and for collecting expert opinions in
158 probabilistic format has been developed for this purpose. The form comprises three main
159 elements for each variable to be elicited: *the question, the information and the demand for*
160 *response*. In the example shown in figure 1, regarding the sliding mechanism, these items are
161 set out in the following way:

- 162 - *The question* corresponds to the variable on which the experts are questioned. It
163 contains a detail on the mechanism to be evaluated, the cross-section to be evaluated
164 (downstream or upstream) and the flood for which the failure probability is demanded.
165 This question on the sliding failure mechanism takes the following form: “*Given all the*
166 *information available, what is the probability P_{fc} that the levee will fail due to sliding of*
167 *the downstream slope in the case of a flood whose level reaches the crest of the levee?*”
168
- 169 - *The information* provides all the data required for the experts to formulate their
170 responses and give their judgements. They correspond to the *objective cue information*
171 presented by [26]. The items of information making up the form are: geometric data,
172 hydraulic data, geotechnical data, the standard cross-section of the profile to be
173 evaluated, the laws of probability of the geotechnical characteristics of the profile to
174 be evaluated and the data resulting from a deterministic analysis represented by the
175 safety factor SF of the levee regarding the slope sliding failure mechanism;
176
- 177 - *The demand for response* corresponds to the way the expert communicates his
178 judgement on the failure probability (P_{fc}) of the cross-section to be evaluated. In view
179 to taking into account the uncertainties linked to levees, the approach proposes expert

180 elicitation in the form of uncertainty interval and a most probable value [1, 13, 27].
 181 Thus the expert is first called on to elicit an uncertainty interval [q5%, q50%, q95%],
 182 then to specify the median value resulting from the interval elicited proposed by [18].
 183 The choice of 5% and 95% quantiles is motivated by the practice of using statistical
 184 evaluations of resistance and stress parameters in civil engineering [8]. Also these are
 185 the quantiles generally used for the application of the classical calibration model [41].
 186

Form 01		Cross-section – Theoretical levee 1	
Question			
Given all the information available, what is the probability P_{fc} that the levee will fail due to sliding of the downstream slope in the case of a flood reaching the height of the levee crest?			
Information			
Geometric data <i>Cf. Figure below</i>	<ul style="list-style-type: none"> - Height = 5 m - Altitude of crest = 205 m NGF - Width of crest = 6 m - Upstream batter = 3/2 (H/V) - Downstream batter = 3/2(H/V) 		
Hydraulic data	<ul style="list-style-type: none"> - Flood level reaching the levee crest - Hydraulic gradient level calculated in steady state 		
Geotechnical data	Body and foundation		
	<ul style="list-style-type: none"> - Sandy-silty material - Permeability $K = 10^{-6} \text{ m/s}$ - Unit weight: $\gamma_h = 18 \text{ kN/m}^3$ 		
Probability laws of input data			
	Angle of friction (body - foundation) Probabilistic law: normal law Mean: 20° Standard deviation: 5°		Cohesion (body - foundation) Probabilistic law: normal law truncated at 0 Mean: 10 kPa Standard deviation: 5 kPa, Min = 0 kPa
Deterministic analysis	Calculated SF (with mean values) Downstream slope 1.3		
Demand for response			
Probabilistic analysis	Probability (P_{fc}) of sliding failure	P_{fc} : quantile q5%	
		P_{fc} : quantile q95%	
		P_{fc} : median value (q50%)	

187
 188 **Figure 1. Example of a sheet for a levee profile resulting from the questionnaire form developed**

189
 190 The questionnaire is given individually to each of the experts. It contains an
 191 introduction briefly presenting the objectives of the study and a short manual for completing

192 the form. Then, each expert freely completes the questionnaire according to his availability
193 within a period of time agreed with the moderator. During this period, each expert may
194 question the moderator if any elements of the questionnaire are unclear. In the same way, each
195 expert may review his answers during this period. A training or dry run with feedback is
196 desirable [19].

197

198 **2.2 Approach for calibrating expert opinions**

199 The approach to calibrating the panel experts' opinions consists in evaluating and weighting
200 the panel experts' opinions in relation to calibration variables whose real values are known.
201 The implementation of the approach for calibrating expert opinions is based on the model
202 developed by Cooke [18]. This model is used to evaluate and weight the experts' opinions
203 with the variables of interest and calibration. The variables of interest are those which the
204 analyst seeks to know the values. The expert judgement is therefore used to build a subjective
205 probabilistic distribution regarding the production of these variables. The calibration variables
206 correspond to variables liable to inform the analyst on the capacity of the experts to provide
207 an accurate and precise estimation. They are of the same nature as the variables of interest.

208 When applying them to levees, it is necessary to define pertinent calibration variables with
209 respect to the variables of interest. Given that the variables of interest are the expert
210 evaluations of risks and of the structural reliability of levees in a probabilistic framework, we
211 adopted them as a calibration variable of the theoretical profiles of levee failure probability
212 with respect to the sliding failure mechanism. This variable can be obtained quantitatively by
213 carrying out a probabilistic reliability analysis, by considering the uncertainties on the input
214 data (cohesion and angle of friction) and by Monte Carlo simulations [42]. The uncertainty of
215 the input data (cohesion and angle of friction) is modelled by laws (see Figure 1). Random
216 draws are made on these values from which stability calculations are made and an evaluation
217 of the probability of failure by Monte Carlo simulation is obtained.

218 This choice of calibration variables has several advantages. The format of the chosen
219 calibration variables is the same as that of the variables of interest: probability of failure of
220 levee profiles. As the calibration variables correspond to the results of a numerical model of
221 theoretical levee profiles, it is possible to construct a base with a large number of calibration
222 variables (using different levee profiles with different geometries, material properties, etc.). It
223 should be noted that the results of a numerical model depend on the assumptions adopted as
224 input data. This allows the elicitation phase of the calibration variables to be easily framed by
225 presenting the input data of the numerical model in the questionnaire.

226 The elicitations performed by the experts can then be compared with the results of the
227 numerical model for the different calibration variables.

228 Implementing Cooke's [18] model permits calculating the individual *performance weight* w_e
229 of each expert, which is also expressed as the *relative calibration weight* w'_e in relation to all
230 the experts of the panel. The calibration weight is determined from the *calibration score* C_e
231 and the *information score* K_e as a function of the following relations [18]:

$$w_e = C_e \cdot K_e \quad [1]$$

$$w'_e = C_e \cdot K_e / \sum_{e=1}^Z C_e \cdot K_e \quad [2]$$

232

where:

- w_e : weight of the expert's calibration e
- w'_e : weight of the relative calibration of expert e in the panel of experts
- C_e : calibration score of expert e
- K_e : information score of expert e
- Z : number of experts

233

234 *Calibration score* C_e permits evaluating the quality (statistical accuracy) of the information
 235 given by the expert. Calibration measures the statistical likelihood that a set of experimental
 236 results corresponds, in a statistical sense, with the experts' assessments. Loosely, the
 237 calibration score is the probability that the divergence between the expert's probabilities and
 238 the observed values of the seed variables might have arisen by chance. A low score (near
 239 zero) means that it is likely, in a statistical sense, that the expert's probabilities are 'wrong'
 240 [19]. The calibration score is evaluated by the following relation 3 below [18]. It is determined
 241 by the addition of an error probability $P(\mathcal{X}^2)$ obtained by the \mathcal{X}^2 distribution for large
 242 samples size (see [41] for more details). We used Kullback-Laibler (KL) divergence measure
 243 which measures the difference between two probability distributions (expected and elicited by
 244 expert), noted $I_e(c_i, p_i)$ within formula 3 and 4 below.

$$C_e = 1 - P(2 * n * I_e(c, p)) = 1 - \mathcal{X}^2(2 * n * I_e(c, p)) \quad [3]$$

245

where :

- $P()$: probability of a random variable according to a \mathcal{X}^2 law
- $\mathcal{X}^2()$: distribution function of a random variable according to the \mathcal{X}^2 law
- n : number of calibration variables
- c : calibration vector representing the portion of true values in each interquartile interval $c = \{c_1, c_2, \dots, c_j\}$
- j : number of interquartile intervals, in this study $j=4$
- p : theoretical probability vector representing the theoretical probabilities of each interquartile interval $p = \{p_1, p_2, \dots, p_j\}$
- $I_e(c_i, p_i)$: relative information between the theoretical probability vector p and the calibration vector c . It is calculated as follows:

246

$$I_e(c, p) = \sum_{i=1}^j c_i \ln(c_i/p_i) \quad [4]$$

247

248 The *information score* K_e permits measuring the quantity of information contained in the
 249 probabilistic distributions given by the experts. It relies on a measure of distances between the

250 vector of subjective probabilities s and the vector of theoretical probabilities p . The
 251 information score is evaluated by the following relation:

252

$$K_e = (1/n) \sum_{i=1}^n I_{e,i}(p, s) \quad [5]$$

253

where:

- s : vector of subjective probability representing the subjective probabilities of each interquartile interval $s = \{s_1, s_2, \dots, s_j\}$. It is used for measuring the distance between the expert's distribution and a non-informative uniform distribution;
- $I_e(p, s)$: the relative information between the vector of subjective probabilities s and the vector of theoretical probabilities p .

254

255 The vector of subjective probability s and the relative information $I_e(p, s)$ are calculated
 256 for each calibration variable by the following formulas:

257

$$s = \left\{ \frac{q_5 - q_0}{q_{100} - q_0}, \frac{q_{50} - q_5}{q_{100} - q_0}, \frac{q_{95} - q_{50}}{q_{100} - q_0}, \frac{q_{100} - q_{95}}{q_{100} - q_0} \right\} \quad [6]$$

258

259

$$I_e(p, s) = \sum_{i=1}^j p_i \ln(p_i/s_i) \quad [7]$$

260

261 2.3 Expert opinion aggregation approach

262 Several approaches have been suggested and/or implemented on case studies for the
 263 aggregation of expert opinions, concerning unweighted simple average aggregations: elicited
 264 probability distributions [14] or elicited quantiles [39], or weighted averages of probabilities
 265 distributions constructed from the elicited quantiles via the respective performances of the
 266 experts on the calibration variables [14]. An interesting discussion of the comparison of the
 267 results of these methods can be found in [29]. The expert opinion aggregation approach is
 268 implemented in this study with performance weighted quantiles [39]. The objective is to
 269 conserve the initial form of the expert elicitations given in probabilistic format with a value
 270 considered as the median and an uncertainty interval. This approach corresponds to a
 271 weighted sum of the calibrated quantiles of expert opinions. The aggregated quantiles
 272 ($q_{a5\%}, q_{a50\%}, q_{a95\%}$) are determined by the following relations:

273

$$q_{a5\%} = \sum w'_e \cdot q_{5\%(e)} \quad [8]$$

$$q_{a50\%} = \sum w'_e \cdot q_{50\%(e)} \quad [9]$$

$$q_{a95\%} = \sum w'_e \cdot q_{95\%(e)} \quad [10]$$

274

where :

- $q_{a5\%}$: aggregated quantile corresponding to a probability of 5%;
- $q_{a50\%}$: aggregated quantile corresponding to a probability of 50%;
- $q_{a95\%}$: aggregated quantile corresponding to a probability of 95%.

275

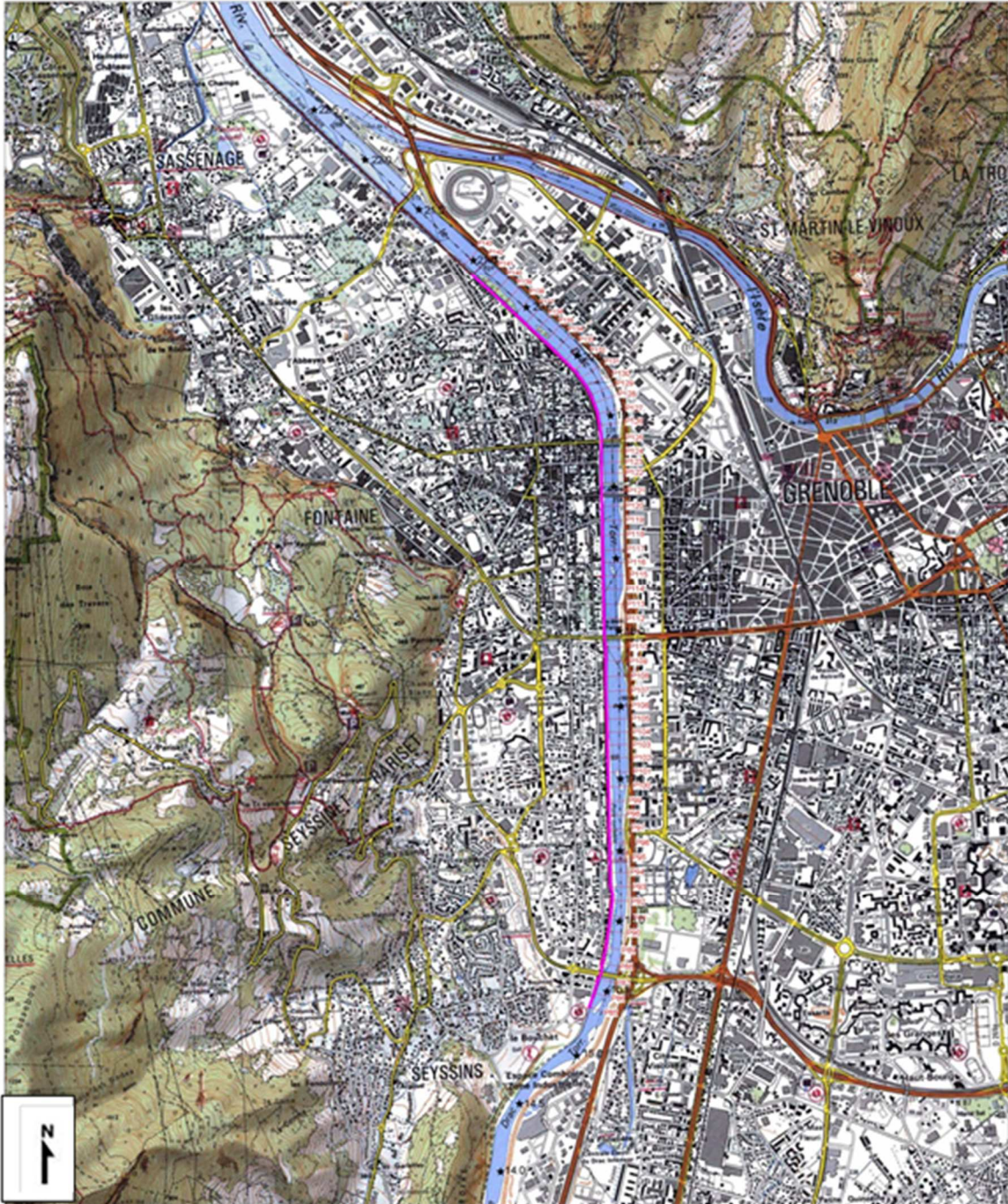
276

277 **3 Application of the approach – Results**

278 **3.1 Presentation of the case study**

279 The levee studied is an earth-fill levee 5500 meters long intended to protect a French city
280 against floods. The height of the levee in the study area varies from 1 m to 6 m.

281



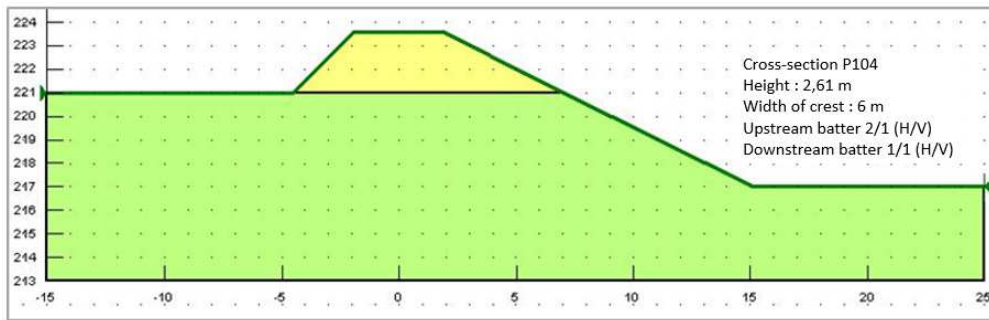
282

283

284

Figure 2. Aerial view of the levee studied (scale: 1/25 000 in A3 format)

285



286

287

Figure 3. Standard cross-section of the levee studied

288

289 The approach developed was implemented by a panel of 6 expert, all civil engineers with
 290 different specialities (geotechnics, river hydraulics, civil engineering structures, risk analysis),
 291 all levee safety experts working for the French Ministry of Ecology, and all volunteers. In
 292 order to realize the experience in the best conditions, the elicitation was done anonymously.
 293 The questionnaire used in this case study contains a brief introduction (including a notice for
 294 completing the form), a section for calibration variables and a section for variables of interest.
 295 This questionnaire was given individually to each of the experts. They provided their
 296 judgements on the calibration variables and the variables of interest corresponding to the
 297 failure probabilities of the levees with respect to the different failure mechanisms. The experts
 298 completed the form within one week. Previously, there was a dry run with a limited number
 299 of variables, which improved the clarity of the form. Other than this dry run, there was no
 300 specific training with feedback on the true values of the calibration variables. This choice was
 301 adopted specifically for this case study for further improvements concerning the experts'
 302 biases.

303 In this study, the calibration variables correspond to the probabilities of sliding failure for 30
 304 theoretical levee cross-sections (cf. section 2.2). This number of calibration variables permits
 305 a robust statistical analysis [43].

306 The study is applied to 40 variables of interest corresponding to 10 levee cross-sections
 307 studied with 4 failure mechanisms per cross-section. The following table summarises the type
 308 of calibration variable and the variables of interest used in this study:

309

310 Table 1. Table summarising the calibration variables and the variables of interest in the application of the approach.

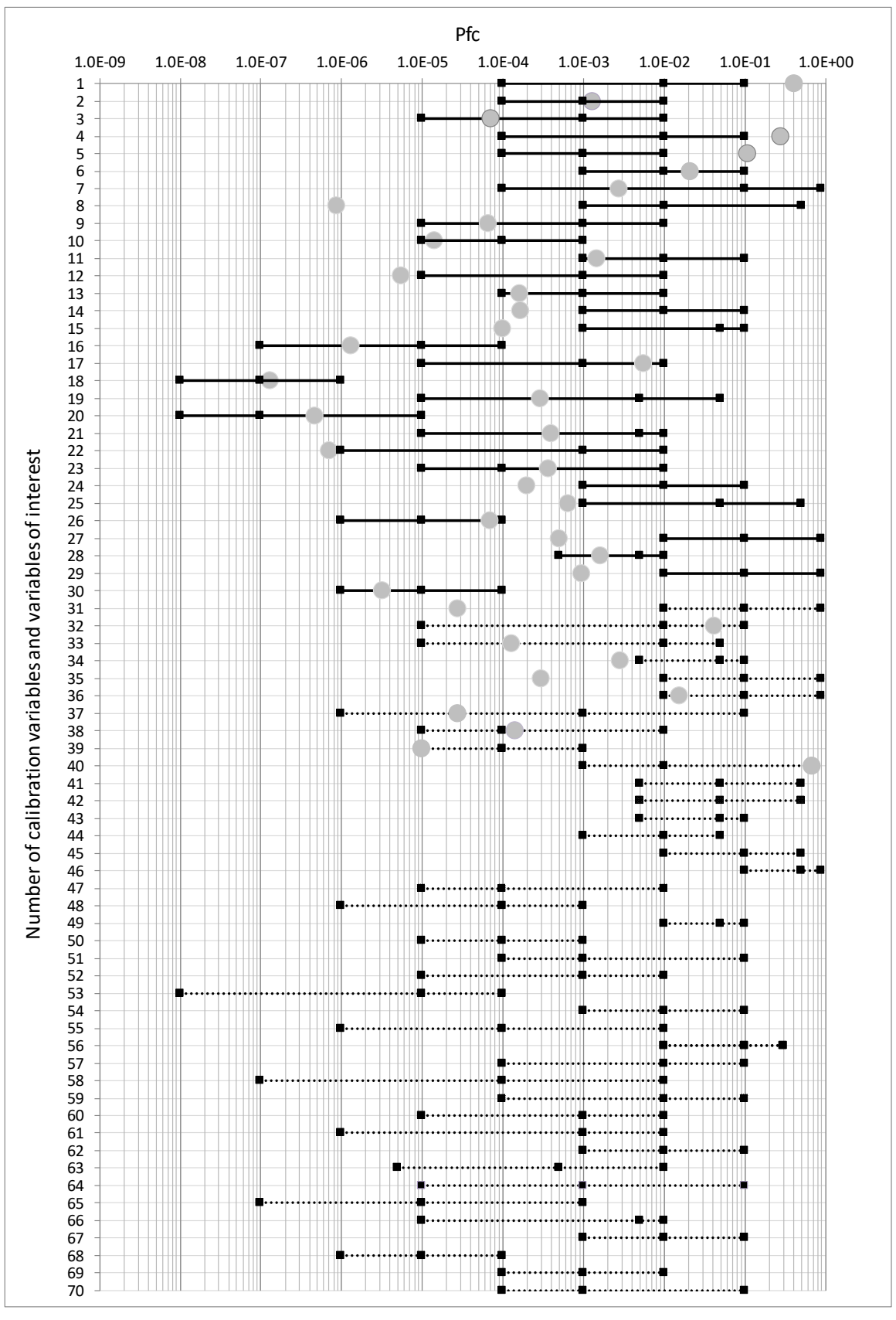
Failure mechanism	Calibration variables (theoretical levee profiles)		Variables of interest (profiles of the levee studied)	
	Number of variables	no. of variable	Number of variables	no. of variable
Sliding	30	no.1 – no.30	10	no.31 – no.40
Internal erosion	-	-	10	no.41 – no.50
Scouring	-	-	10	no.51 – no.60

Overtopping	-	-	10	no.61 – no.70
-------------	---	---	----	---------------

311

312 **3.2 Results of the elicitation phase**

313 The following graph presents the elicitations of expert no.1 regarding the failure probabilities
314 (P_{fc}) of 4 failure mechanisms. The black squares correspond to the quantiles (5%, 50% and
315 95%) elicited by the expert, while the grey circles correspond to the “true value” of the
316 calibration variable (evaluated through a reliability analysis).



318
319
320

Figure 4. Raw elicitations of expert no.1 relating to calibration variables no.1 to no.30 and to variables of interest no.31 to no.40 – sliding failure mechanism, no.41 to no.50 – internal erosion failure mechanism, no.51 to no.60 – scouring failure mechanism), no.61to no.70 – Overtopping failure mechanism.

321 Table 2 gives the distribution of the true values (in number and in percentage) in the
322 interquartile intervals elicited by the experts for theoretical levee cross-sections no.1 to no.30:

323

324
325

Table 2. Distribution of true values of the calibration variables (no.1-no.30) in the interquartile intervals elicited by the experts and the mean of elicitations.

Calibration variable (no.1 to no.30)					
Interquartile interval		[0%, 5%]	[5%, 50%]	[50%, 95%]	[95%, 100%]
Expert no.1					
(P_{fc}) (no.1 to no.30)	n	9	11	7	3
	%	30%	37%	23%	10%
Expert no.2					
(P_{fc}) (no.1 to no.30)	n	9	10	10	1
	%	30%	33%	33%	3%
Expert no.3					
(P_{fc}) (no.1 to no.30)	n	10	5	7	8
	%	33%	17%	23%	27%
Expert no.4					
(P_{fc}) (no.1 to no.30)	n	10	8	8	4
	%	33%	27%	27%	13%
Expert no.5					
(P_{fc}) (no.1 to no.30)	n	0	4	15	11
	%	0%	13%	50%	37%
Expert no.6					
(P_{fc}) (no.1 to no.30)	n	8	8	6	8
	%	27%	27%	20%	27%
Means of elicitations (before calibration)					
(P_{fc}) (no.1 to no.30)	n	7.67	7.67	8.83	5.83
	%	26%	26%	29%	19%

326

327 The results show that the percentage of true values in the interval [5%, 95%] varies from 40%
328 (17%+23%) for expert no.3to 66% (33%+33%) for expert no.2. More exactly for expert 3,
329 33% in the interval [0%, 5%] instead of 5%, 17% in the interval [5%; 50%] instead of 45%,
330 23% in the interval [50%, 95%] instead of 45, and 27% in the interval [95%, 100%] instead of
331 5%, and for expert 2 : 30% in the interval [0%, 5%] instead of 5%, 33% in the interval [5%;
332 50%] instead of 45%, 33% in the interval [50%, 95%] instead of 45, and 3% in the interval
333 [95%, 100%] instead of 5%. For experts nos. 1, 4 and 6, the true values are concentrated in
334 the interval [0%, 50%], expressing a trend towards overestimating the failure probabilities. On

335 the contrary, for expert no.5, the true values are concentrated in the interval [50%, 100%],
 336 expressing a trend towards underestimating the failure probabilities.

337 3.3 Results of the calibration phase

338 The calibration phase is performed using 30 calibration variables ($n=30$). The result of the
 339 calibrating the expert opinions is presented as a calibration score, information score and
 340 relative calibration weight.

341 Since elicitation is performed on the $q_{5\%}$, $q_{50\%}$ and $q_{95\%}$ quartiles, the theoretical probability
 342 vector p associated with the interquartile intervals is $p=\{0.05, 0.45, 0.45, 0.05\}$. The
 343 calibration vectors c can be obtained directly from table 2 (for expert 1, $c=\{0.30, 0.37, 0.23,$
 344 $0.10\}$). Then, the relative information between c and p vectors, $I_e(c,p)$ is calculated according
 345 Eq.4 (for expert 1, $I_e(c,p)=0.38$), which finally gives the calibration score C_e presented in
 346 Table 3 by using Eq.5.

347 For the calculation of the information score, the elicited failure probability values can be very
 348 low, expressed in the format (10^{-x}) . We change the variable in order to conserve a constant
 349 difference in absolute value between (10^{-x}) and $(10^{-(x-1)})$, which amounts to applying the
 350 common logarithm to the elicited probabilities ($\log_{10}(P_{fc})$). This change of variable permits a
 351 more pertinent interpretation of the information score. The subjective probability vector s is
 352 then evaluated for each expert and for each of the calibration variables. For example, expert 1
 353 elicits the probability values of 10^{-4} , 10^{-2} and 10^{-1} for the quartiles 5%, 50% and 95%
 354 respectively for the first calibration variable. With the variable change to base \log_{10} , we thus
 355 obtain $q_{5\%}=-4$, $q_{50\%}=-2$ and $q_{95\%}=-1$, which makes it possible to evaluate the vector $s_1=\{0.24,$
 356 $0.39, 0.19, 0.18\}$ with Eq.6 and the relative information $I_{e,1}(p,s)=0.31$ with Eq.7. The
 357 information score K_e is finally evaluated with Eq.5, and corresponds to the average of the
 358 relative information $I_{e,i}(p,s)$ evaluated for the 30 calibration variables.

359 Once the calibration score C_e and information score K_e have been evaluated, the individual
 360 weight w_e and the relative weight of calibration w'_e are directly evaluated by Eq.1 and Eq.2
 361 respectively. The results of the calibration score C_e , information score K_e , individual weight
 362 w_e , and relative weight of calibration w'_e are presented in the following table for each expert.
 363 The calculations to obtain these results were carried out in an Excel file.

364

365 Table 3. Results of the expert opinion calibration phase in relation to the calibration variables.

	Calibration score	Information score	Weight of individual	Relative weight
No. expert	C_e	K_e	w_e	w'_e
1	0.79	0.74	0.59	0.28
2	0.91	0.91	0.83	0.40
3	0.03	0.79	0.02	0.01
4	0.46	1.04	0.48	0.23
5	0.14	0.44	0.06	0.03
6	0.19	0.52	0.10	0.05
			Total	1

366

367 The results obtained from the calibration of expert opinions showed that the values of the
368 calibration scores (C_e) varied from 3% to 91%. The highest value was obtained by expert no.
369 2, meaning that their elicitations contained more true calibration values in comparison to the
370 elicitations of the other panel experts. On the contrary, the lowest calibration score was
371 obtained by expert no. 3, meaning that their elicitations contained the fewest true calibration
372 values.

373 Regarding the information score, the values varied from 0.44 to 1.04. The highest value was
374 obtained by expert no.4, meaning that the uncertainty intervals elicited by this expert were
375 narrower than those given by the other panel experts. On the contrary, the lowest value of this
376 score was obtained by expert no.5, meaning that the uncertainty intervals elicited by this
377 expert were wider than those given by the other panel experts.

378 We observed that the calibration score in our study had a more significant influence than the
379 information score on the final result of the expert opinion calibration phase. Indeed, the range
380 of the final calibration score was wider than that of the information score. Thus, expert no.2
381 with the highest calibration score obtained the best relative calibration weight ($w'_2 = 40\%$)
382 whereas expert no.3 with the lowest calibration score obtained the lowest relative calibration
383 weight ($w'_3 = 1\%$). The question of not taking into account his opinion may then arise. In the
384 context of this study, we preferred to keep all the opinions, taking into account the fact that
385 that of some experts does not significantly influence the final results.

386

387 *3.4 Results of the aggregation phase*

388 The result of the expert opinion aggregation phase is shown by one probabilistic distribution
389 relative to each levee cross-section evaluated by the experts. We operated results exposed in
390 table 3 to combine expert assessments, using quantiles combination [39] to produce results
391 exposed in figure 5 and table 4. The following graph shows the result of the aggregation of
392 the calibrated expert opinions for the 30 calibration variables related to the theoretical levees
393 (continuous lines) where the true values are represented by a circle. The calculations to obtain
394 these results were carried out in an Excel file.

395

396

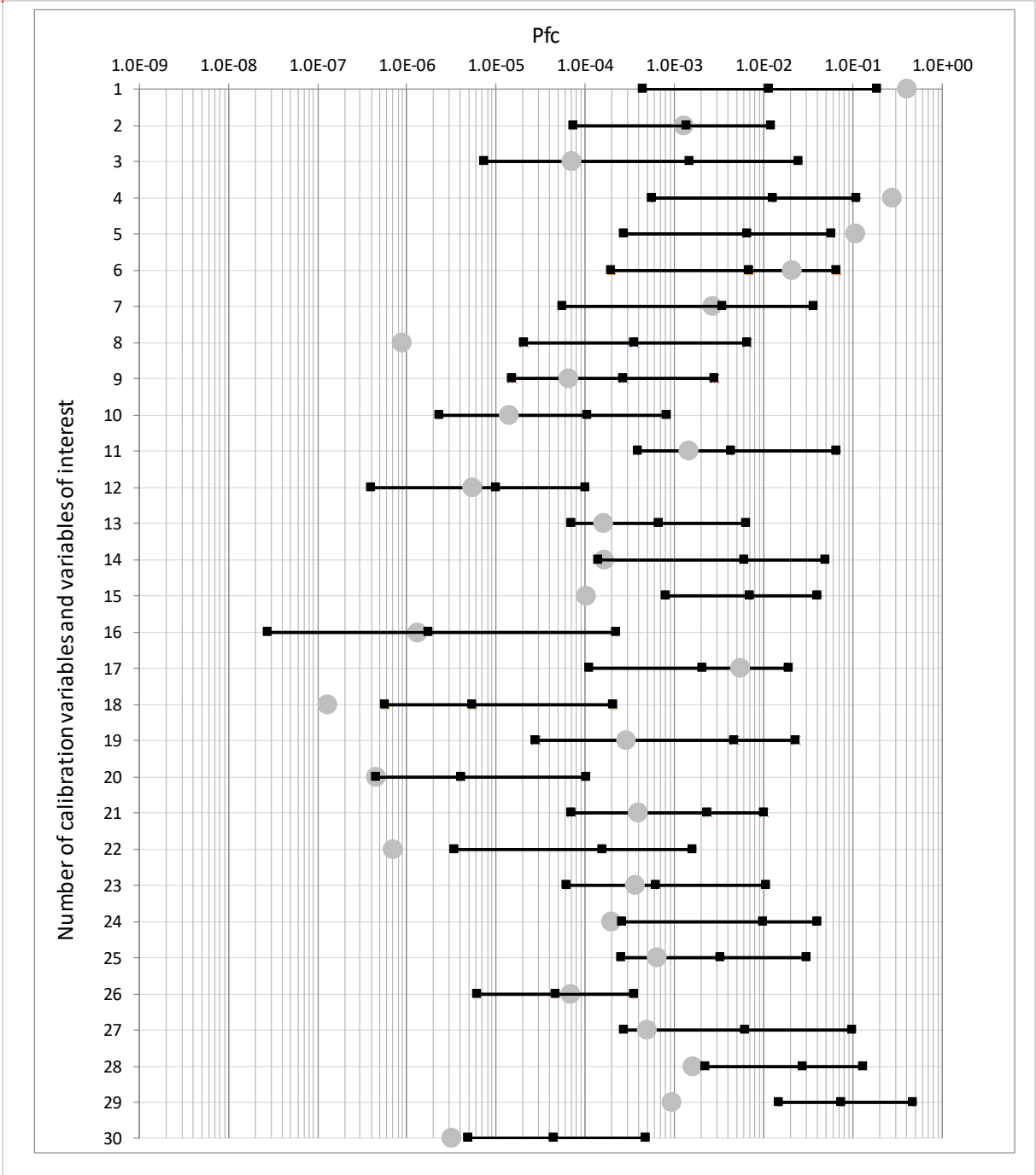


Figure 5. Calibration variables calibrated and aggregated and their uncertainty intervals

397
 398
 399
 400
 401
 402
 403
 404

Figure 5 can be interpreted using Table 4 below which presents the true calibration variables in the interquartile intervals at the end of the expert opinion aggregation phase:

405

Table 4. Distribution of the true calibration variable values (no.1-no.30) in the interquartile intervals.

Calibrated and aggregated elicitations of calibration variables (no.1 to no.30)					
Interquartile interval		[0%, 5%]	[5%, 50%]	[50%, 95%]	[95%, 100%]
(P_{fc}) (no.1 to no.30)	n	9	15	3	3
	%	30%	50%	10%	10%

406

407 At the end of the aggregation phase we observed that the final distributions of the calibrated
 408 and aggregated expert opinions contained true values in the uncertainty interval for 18
 409 calibration variables out of a total of 30 variables, i.e. 60% of the true values belonging to the
 410 uncertainty interval [5%, 95%]. Thus, the aggregation of the calibrated expert opinions leads
 411 to an uncertainty interval [5%, 95%] containing more true values in comparison to the raw
 412 expert uncertainty intervals.

413

414 3.5 Application to variables of interest

415 On the basis of the expert calibrations, the approach was applied to the variables of interest.
 416 Figure 6 presents the results for 40 variables of interest corresponding to 10 levee cross-
 417 sections studied for 4 failure mechanisms:

- 418 - variables no.31 to no.40 corresponding to the sliding failure mechanism,
- 419 - variables no.41 to no.50 corresponding to the internal erosion failure mechanism,
- 420 - variables no.51 to no.60 corresponding to the scouring failure mechanism,
- 421 - variables no.61 to no.70 corresponding to the overtopping failure mechanism.

422

423 Table 5 shows the distribution of true values (in number and in percentage) in the raw
 424 interquartile intervals (before calibration and aggregation) elicited by the experts for the
 425 sliding failure mechanism for levee cross-sections studied no.31 to no. 40:

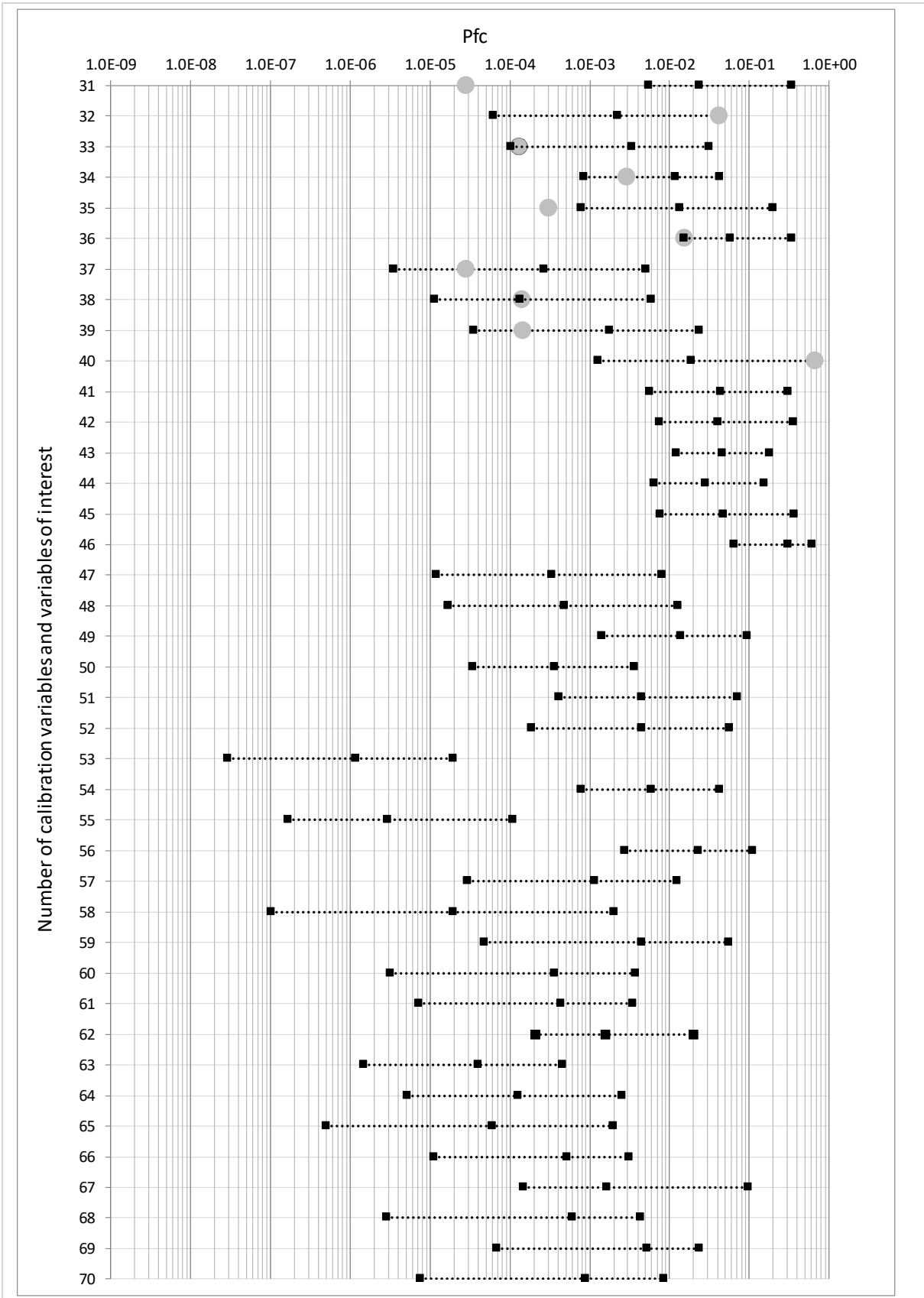
426

427 Table 5. Distribution of true values of variables of interest (no.31-no.40) in the interquartile intervals elicited by the
 428 experts for the sliding failure mechanism.

Variable of interest (no.31 to no.40) for the sliding failure mechanism					
Interquartile interval		[0%, 5%]	[5%, 50%]	[50%, 95%]	[95%, 100%]
Expert no.1					
(P_{fc}) (no.31 to no.40)	n	3	3	4	0
	%	30%	30%	40%	0%
Expert no.2					
(P_{fc}) (no.31 to no.40)	n	4	3	3	0
	%	40%	30%	30%	0%

Expert no.3					
(P_{fc}) (no.31 to no.40)	n	3	2	4	1
	%	30%	20%	40%	10%
Expert no.4					
(P_{fc}) (no.31 to no.40)	n	3	5	2	0
	%	30%	50%	20%	0%
Expert no.5					
(P_{fc}) (no.31 to no.40))	n	0	0	6	4
	%	0%	0%	60%	40%
Expert no.6					
(P_{fc}) (no.31 to no.40)	n	0	1	5	4
	%	0%	10%	50%	40%
Mean of elicitations (before calibration)					
(P_{fc}) (no.31 to no.40)		2.17	2.33	4	1.5
		21.7%	23.3%	40%	15%

429



430

431
432
433

Figure 6. Calibrated and aggregated variables of interest and their uncertainty intervals: no.31 to no.40 – sliding failure mechanism, no.41 to no.50 – internal erosion failure mechanism, no.51 to no.60 – scouring failure mechanism, no.61 to no.70 – overtopping failure mechanism.

434 Regarding the different variables of interest considered in this study, it is possible to perform
 435 a quantitative evaluation by analysing the failure probabilities for variables no.31 to no.40 of
 436 the sliding failure mechanism, which correspond to the sliding failure probabilities of six
 437 cross-sections of the levee studied. These variables allow comparing the results of the
 438 calibration and aggregation approach applied to the expert opinions with the results of a
 439 quantitative analysis. Table 6 below presents the distribution of true values in the interquartile
 440 intervals for the variables of interest no.31 to no.40 at the end of the expert opinion
 441 aggregation phase and using the calibration weights determined during the expert opinion
 442 calibration phase for the cross-sections of the theoretical levees:

443

444 **Table 6. Distribution of true values of the variables of interest (no.31-no.40) in the interquartile intervals.**

Calibrated and aggregated elicitation pf variables of interest (variables no.25-no.30)					
Interquartile interval		[0%, 5%]	[5%, 50%]	[50%, 95%]	[95%, 100%]
(P_{fc}) (no.31- no.40)	n	2	5	2	1
	%	20%	50%	20%	10%

445

446 The final distributions of the aggregated expert opinions of variables no.31 to no.40 contain
 447 the true value in the uncertainty interval [5%, 95%] seven times, giving a percentage of 70%
 448 of the true values of the uncertainty values given by the experts. The evaluation of variables
 449 31 and 40 is similar to a cross-validation of the results obtained for the calibration variables (1
 450 to 30). We find that we get similar results. We propose to discuss this in paragraph 3 of
 451 section 4.2.

452 Regarding the other failure mechanisms (internal erosion, overtopping and scouring), the
 453 structural limit states were not quantifiable since it was not possible to quantitatively evaluate
 454 the failure probability of these failure mechanisms. The true values were therefore unknown
 455 for these mechanisms and the approach proposed therefore made it possible to implement the
 456 expert judgement to evaluate these failure probabilities in the framework of a levee risk
 457 analysis.

458 **4 Discussion**

459 The approach developed comprises a protocol for using expert judgement in the area of river
 460 levees to evaluate their structural reliability regarding the failure probabilities of several
 461 failure mechanisms. Our discussion focuses on the approach developed and on the results
 462 obtained.

463 **4.1 Discussion on the approach developed**

464 The approach proposed answers the issue of evaluating the reliability of levees as raised by
 465 French regulations relating to hazard studies and determining the probability of failure on the
 466 basis of expert evaluations. The approach permits eliciting, calibrating and aggregating the

467 evaluations of experts in different disciplines. The treatments of the expert elicitations
468 allowed us to obtain an evaluation of failure probability and an evaluation of the uncertainty
469 on this probability by way of an uncertainty interval. The application of this methodological
470 approach led to two significant improvements: i) it combined additional expert opinions from
471 the calibration and information scores to give more weight to the most relevant experts; ii) it
472 used an uncertainty interval to provide quantitative information on the uncertainty on the final
473 failure probability. In addition, the analysis of the uncertainty interval can be completed with
474 the information and calibration scores obtained for the different experts, or, more intuitively,
475 by visualising the quantities of calibration variables falling outside the uncertainty intervals
476 given by the experts.

477 The major difficulty for implementing the model concerns the construction of the calibration
478 variables. In our work, we chose the sliding failure scenario as the calibration variable of a
479 levee failure. The calibration variables were obtained by analysing probabilistic reliability by
480 taking into account the uncertainties on the input data. However, as we mentioned in
481 introduction within our research study, there was no calibration variable directly involving the
482 failure probability of a levee for internal erosion, overtopping and scouring failure scenarios.
483 Also, the use of calibration variables focused on the sliding mechanism to assign calibration
484 weights to the other mechanisms (internal erosion, overtopping and scouring) could be
485 discussed on the pertinence of the process.

486 However, we consider that our approach is relevant because the parameters involved in the
487 mechanics of internal erosion, overtopping and scouring are the geotechnical parameters of
488 the soils and of the levee also involved in the sliding mechanism: the density of the materials,
489 cohesion, angle of friction, geometry of the levee, etc. Thus, calibrating the experts on the
490 sliding mechanism amounts to calibrate the experts on the geotechnical parameters involved
491 in the sliding mechanism, and consequently on the other mechanisms of failure. Therefore this
492 approach of calibration of the variables of interest seems well founded and justified.

493 One perspective for improving the calibration variables of the latter scenarios would be to
494 break these failure scenarios down into elementary functional events [17, 44], for which it
495 would be easier to obtain calibration variables. It would then be possible to question the
496 experts on the calibration variables linked to elementary events and obtain the failure
497 probability for these scenarios by multiplying the probabilities elicited for the elementary
498 variables.

499 *4.2 Discussion on the results*

500 It is considered that there is an overconfidence bias when the intervals estimated by the
501 experts are too narrow [45, 25]. We propose in this study to consider two sub-components of
502 the over-confidence bias, excess of optimism and excess of caution. Situations of bias on
503 excess of optimism and excess of caution can be identified when the true values of the
504 calibration variables fall within the intervals [95%; 100%] and [0%; 5%] in more than 10% of
505 the expert's or the experts' responses. If the value of the calibration variable is in the [0%;
506 5%] interval, this means that the expert has given a higher failure probability than that of the
507 true value of the calibration variable, in which case the evaluation can be deemed as over-
508 cautious, possibly leading to costly structural reinforcement works. On the contrary, if the true
509 value of the calibration variable falls within the [95%; 100%] interval, this means that the

510 expert was overconfident, since they gave a lower failure probability and uncertainty interval
511 than the real failure probability of the structure. This situation is dangerous because it may
512 lead to a poor evaluation of the structure's failure probability liable to place populations in
513 danger.

514 The results obtained concerning the elicitations of experts before treatment confirm previous
515 observations of the probabilistic evaluation based on expert opinion and allow distinguishing
516 biases of excess of caution and excess of optimism. We observed that the experts tended to be
517 overcautious regarding the evaluation of failure probabilities and to be overconfident
518 regarding the evaluation of uncertainties. Indeed, the results presented in table 2 show that on
519 average 26% of the expert evaluations were cautious, instead of the 5% that should have been
520 obtained if the distribution of the experts were perfectly calibrated. As mentioned above, this
521 bias of caution points towards safety in comparison to a bias of excess of optimism. The bias
522 of excess of optimism obtained reached 19%, which is high for structures critical for the
523 safety of people and property. It is noteworthy that the literature concerning heuristic biases
524 agrees in saying that they can be reduced by implementing several procedures: in the
525 calibration phase as presented in this article, or in the elicitation phase by making the experts
526 aware of biases and their explicit description [12], or in the framework of calibration exercises
527 prior to elicitation. This provides an interesting perspective of improvement for this study.

528 The results obtained by applying our calibration approach to calibration variables (nos.1 to
529 30) showed a marked reduction of variables in the excess of optimism interval [95%; 100 %],
530 thus falling from 19% (mean of individual expert elicitations) to 10% (table 4), which is a
531 considerable improvement that goes towards ameliorating the safety of the structures
532 concerned. There is still room for progress to reach the target value of 5% in the excess of
533 optimism interval [95%; 100%]. The results obtained for the variables of interest (no.31 and
534 40) are analogous with a substantial reduction of variables in the excess of optimism interval
535 [95%; 100 %] reduced from 15% to 10% (cf. table 5 and table 6). To improve this result, it
536 would be interesting to implement specific mathematical procedures to correct these types of
537 bias using inflation factors [46].

538

539 **5 Conclusion**

540 Studying the structural reliability of levees with quantitative approaches is complex and often
541 impossible due to the lack of statistical data or the lack of limit state equations for certain
542 mechanisms. Thus this article presented an approach for evaluating the failure probabilities of
543 levees based on the calibration of expert judgement. The aim of the procedure was to
544 determine a calibration weight that can be used for other levee risk analyses.

545 The approach developed is pertinent for using expert judgement in the area of river levees. It
546 can be used to evaluate the reliability of structures in terms of failure probability based on this
547 judgement. The approach first proposes eliciting expert opinions in probabilistic format then
548 identifying a calibration weight for the best expert elicitation in view to aggregating them. At
549 the end of an analysis of the accuracy and informativeness of expert opinions, it assigns a high

550 calibration weight to the most accurate expert elicitations and a low calibration weight to
551 inaccurate and thus less informative elicitations.

552 The results obtained from the application of the approach developed in the area of levees
553 showed the advantage of calibrating expert judgement. Finally, the approach can be used to
554 correct through aggregation expert evaluations of levee reliability and obtain a more
555 informative aggregated evaluation.

556 The calibration variables used in the case study correspond to probabilities of sliding failure
557 evaluated by a reliability analysis. The accuracy of the expert judgments calibration can be
558 assessed for the sliding mechanism by comparison with the results of a reliability analysis.
559 However, the accuracy of the calibration cannot be evaluated for the other failure mechanisms
560 that have not been assessed by a reliability analysis in this case study. Some research works
561 are available in the literature concerning the reliability analysis of levees with respect to other
562 failure mechanisms such as internal erosion or pipping [5]. As these works have not been
563 considered herein, one perspective should be to decline the methodology of this paper by also
564 using calibration variables corresponding to other levee failure mechanisms (such as the
565 probability of failure due to internal erosion assessed by a reliability analysis). This would
566 allow the results to be compared in order to further explore the validation/verification of
567 calibration of experts across different failure modes.

568 Besides, the approach has limitations regarding the quantitative correction of the trends of
569 overconfidence followed by expert opinions. Another perspective for this work would be to
570 improve correction with correction protocols including the treatment of expert biases, such as
571 overconfidence which is one of the most common (and potentially severe) problem in expert
572 judgment [45]. In this study, each expert is questioned individually, without allowing
573 discussion or consensus among experts (as in the case of collective elicitation by a panel of
574 experts). Future work could study and compare these approaches (individual and collective
575 elicitation) for application to flood protection levees.

576 This approach developed for river levees could be adapted and applied in other areas where
577 recourse to expert judgement is the only means of obtaining information usable for carrying
578 out a reliability analysis. The approach will require adjustments specific to each category of
579 civil engineering structure before its implementation.

580 **6 References**

581 [1] Mériaux P., Royet P., Folton C. (2001). *Surveillance, entretien et diagnostic des digues de*
582 *protection contre les inondations*. Cemagref Editions, 191p.
583 [https://www.eyrolles.com/BTP/Livre/surveillance-entretien-et-diagnostic-des-digues-de-](https://www.eyrolles.com/BTP/Livre/surveillance-entretien-et-diagnostic-des-digues-de-protection-contre-les-inondations-9782853626361/)
584 [protection-contre-les-inondations-9782853626361/](https://www.eyrolles.com/BTP/Livre/surveillance-entretien-et-diagnostic-des-digues-de-protection-contre-les-inondations-9782853626361/)

585 [2] Ciria, MEDDE (Ministère de l'Ecologie du Développement durable et de l'Energie),
586 USACE (US Army Corps of Engineers). (2013). *The International Levee Handbook*.

- 587 CIRIA, London. (ISBN: 978-0-86017-734-0).
588 https://www.ciria.org/Resources/Free_publications/ILH.aspx
- 589 [3] Vrijling J.K., (2001). Probabilistic design of water defense systems in The Netherlands.
590 *Reliability Engineering and System Safety*, 74, 337–344. [https://doi.org/10.1016/S0951-](https://doi.org/10.1016/S0951-8320(01)00082-5)
591 [8320\(01\)00082-5](https://doi.org/10.1016/S0951-8320(01)00082-5)
- 592 [4] Apel H., Annegret H. Thieken A. H., Merz B., Blöschl G. (2006). A probabilistic
593 modelling system for assessing flood risks. *Natural Hazards*, 38 (1–2), 79–100.
594 <https://doi.org/10.1007/s11069-005-8603-7>
- 595 [5] Mazzoleni M., Barontini S., Ranzi R., Brandimarte L. (2014). An innovative probabilistic
596 methodology for evaluating the reliability of discrete levee reaches owing to piping.
597 *Journal of Hydrologic Engineering*, 20, 04014067.
598 <https://ascelibrary.org/doi/10.1061/%28ASCE%29HE.1943-5584.0001055>
- 599 [6] Dawson R., Hall J., Sayers P., Bates P., Rosu C. (2005). Sampling-based flood risk
600 analysis for fluvial dike systems, *Stochastic Environmental Research and Risk*
601 *Assessment*, 19, 388–402. <https://link.springer.com/article/10.1007/s00477-005-0010-9>
- 602 [7] Vorogushyn S., Merz B., Lindenschmidt K-E., Apel H. (2010). A new methodology for
603 flood hazard assessment considering dike breaches. *Water Resources Research*, 46, 1–17.
604 Doi: 10.1029/2009WR008475
- 605 [8] Peyras L., Merckle S., Royet P., Bacconnet C., Ducroux A. (2010). Study on a semi-
606 probabilistic method for embankment hydraulic works – Application to sliding
607 mechanism. *European Journal of Environmental and Civil Engineering*, Vol. 14, no.5.
608 Pp.669-691 <https://doi.org/10.1080/19648189.2010.9693253>
- 609 [9] Peyras L., Carvajal C., Felix H., Bacconnet C., Royet P., Becue JP., Boissier D. (2012).
610 Probability-based assessment of dam safety using combined risk analysis and reliability
611 methods – application to hazards studies, *European Journal of Environmental and Civil*
612 *Engineering*, <https://doi.org/10.1080/19648189.2012.672200>
- 613 [10] Perlea M., Ketchum E. (2011). Impact of Non-Analytical Factors in Geotechnical Risk
614 Assessment of Levees. ASCE Georisk, June 26-28 2011, Atlanta, Georgia, USA. DOI:
615 10.1061/41183(418)117 <https://ascelibrary.org/doi/10.1061/41183%28418%29117>
- 616 [11] Serre D., Peyras L., Tourment R., Diab Y. (2008). Levee Performance Assessment
617 Methods Integrated in a GIS to Support Planning Maintenance Actions, *ASCE - Journal*
618 *of Infrastructure Systems*, 10.1061/(ASCE)1076-0342(2008)14:3(201), 201-213.
619 [https://ascelibrary.org/doi/10.1061/%28ASCE%291076-](https://ascelibrary.org/doi/10.1061/%28ASCE%291076-0342%282008%2914%3A3%28201%29)
620 [0342%282008%2914%3A3%28201%29](https://ascelibrary.org/doi/10.1061/%28ASCE%291076-0342%282008%2914%3A3%28201%29)
- 621 [12] Vuillet M., Peyras L., Carvajal C., Serre D., Diab Y. (2013). Levees performance
622 evaluation based on subjective probability *European Journal of Environment and Civil*
623 *Engineering*, Volume 17, issue 5/2013, pp 329-349.
624 <https://doi.org/10.1080/19648189.2013.785723>

- 625 [13] Vuillet M., Peyras L., Carvajal C., Diab Y., (2016). Developing a Probabilistic Multi-
626 criteria Method for River Levee Performance Evaluation in Support to Diagnostic
627 Analysis, *ASCE- Journal of Infrastructures System*, 04016008-1, 9p.
628 <https://ascelibrary.org/doi/10.1061/%28ASCE%29IS.1943-555X.0000283>
- 629 [14] O'Hagan A., Buck C.E., Daneshkhan A., Eiser J.R., Garthwaite P.H., Jenkinson D.J.,
630 Oakley J.E., Rakow T., (2006). *Uncertain Judgments: Eliciting Experts Probabilities*,
631 West Sussex, UK, John Wiley&Sons Inc. ISBN: 978-0-470-02999-2. 338p.
632 [https://www.wiley.com/en-fr/Uncertain+Judgements:+Eliciting+Experts'+Probabilities-
633 p-9780470029992](https://www.wiley.com/en-fr/Uncertain+Judgements:+Eliciting+Experts'+Probabilities-+p-9780470029992)
- 634 [15] Cooke R.M., Goossens L.H.J. (2008). TU Delft expert judgment data base. *Reliability
635 Engineering and System Safety* 93 657 – 674 <https://doi.org/10.1016/j.ress.2007.03.005>
- 636 [16] Jaiswal K.S., Wald D.J., Perkins D., Aspinall W.P. and Kiremidjian A.S. (2014).
637 Estimating structural collapse fragility of generic building typologies using expert
638 judgment. Chap 117 in: *Safety, Reliability, Risk and Life-Cycle Performance of
639 Structures and Infrastructures* (eds: Deodatis, G., Ellingwood, B.R., Frangopol, D.M.),
640 CRC Press; 879-886. <https://doi.org/10.1201/b16387-130>
- 641 [17] Vick S G. (2002). *Degrees of Belief: Subjective Probability and Engineering Judgment*.
642 *American Society of Civil Engineers*. United States of America, 455p.
643 <https://cedb.asce.org/CEDBsearch/record.jsp?dockkey=0130899>
- 644 [18] Cooke R. (1991). *Experts in uncertainty: opinion and subjective probability in science*.
645 Oxford University Press, 336 p.
- 646 [19] Cooke R.M., Goossens L.H.J. (2004). Expert judgement elicitation for risk assessment of
647 critical infrastructures. *Journal of Risk Research*; 7(6):643–56.
648 <https://doi.org/10.1080/1366987042000192237>
- 649 [20] Meyer M.A., Booker J.M. (1993). Eliciting and Analyzing Expert Judgment: A practical
650 Guide, *ASA-SIAM Series on Statistics and Applied Probability*.
651 <https://doi.org/10.1137/1.9780898718485>
- 652 [21] Morgan M.G., Henrion M. (1990). *Uncertainty - A Guide to Dealing with Uncertainty in
653 Quantitative Risk and Policy Analysis*, Cambridge, Royaume-Uni, Cambridge University
654 Press.[https://doi.org/10.1002/\(SICI\)1099-0771\(199606\)9:2<147::AID-
655 BDM199>3.0.CO;2-8](https://doi.org/10.1002/(SICI)1099-0771(199606)9:2<147::AID-BDM199>3.0.CO;2-8)
- 656 [22] Clemen R.T., Reilly T. (2001). *Making hard decisions with decision tools*. Duxbury
657 Press, Pacific Grove, CA, 733p.
- 658 [23] Ayyub B. (2001). *Elicitation of expert opinions for uncertainty and risks*. CRC press.
659 ISBN 9780849310874, 328p.

- 660 [24] Garthwaite, P.H., Kadane, J.B., O'Hagan, A. (2005). Statistical methods for eliciting
661 probability distributions. *Journal of the American Statistical Association*, 100, 680-701.
662 <https://doi.org/10.1198/016214505000000105>
- 663 [25] Burgman, M. (2015). *Trusting Judgements: How to Get the Best out of Experts*.
664 Cambridge: Cambridge University Press 203p.
- 665 [26] Soll J. B. (1996). Determinants of overconfidence and miscalibration: The roles of
666 random error and ecological structure. *Organizational Behavior and Human Decision*
667 *Processes*, 65, 117-137. <https://doi.org/10.1006/obhd.1996.0011>
- 668 [27] Soll J.B., Klayman J. (2003). Overconfidence in intervals estimates. *Journal of*
669 *Experimental Psychology: Learning, memory and cognition*, 30, 299-314.
670 <http://dx.doi.org/10.1037/0278-7393.30.2.299>
- 671 [28] Mason A.J., Gomes M., Grieve R., Carpenter J.R. (2018). A Bayesian framework for
672 health economic evaluation in studies with missing data. *Health Economics*. Wiley &
673 Sons, 2018; 1-14. <https://doi.org/10.1002/hec.3793>
- 674 [29] Colson A.R., Cooke R.M. (2018). Expert Elicitation: Using the Classical Model to
675 Validate Experts' Judgments. In *Review of Environmental Economics and Policy*,
676 Volume 12, Issue 1, 1 February 2018, Pages 113–132,
677 <https://doi.org/10.1093/reep/rex022>
- 678 [30] Martire K.A., Grows B., Navarro D.J. (2018). What do the experts know? Calibration,
679 precision, and the wisdom of crowds among forensic handwriting experts. *Psychonomic*
680 *Bulletin & Review* December 2018, Volume 25, Issue 6, pp 2346-2355
681 <https://doi.org/10.3758/s13423-018-1448-3>
- 682 [31] Cooke R.M., Jager E. (1998). Failure Frequency of Underground Gas Pipelines, *Risk*
683 *Analysis*, vol. 1, no 4, 511-527, 1998. 10.1111/j.1539-6924.1998.tb00365.x
- 684 [32] Cooke R.M., Jager E., Lewandowski D. (2002). "Reliability model for underground gas
685 pipelines" Probabilistic Safety Assessment and Management E.J. Bonano, A.L. Camp,
686 M.J. Majors, R.A. Thompson (eds), Elsevier, 2002; 1045-1050.
687 <https://fr.scribd.com/document/143758470/reliability-gas-pipelines>
- 688 [33] Cooke R.M., Jager E., Lewandowski D. (2003). Reliability Model for Underground Gas
689 Pipelines Case Studies in *Reliability and Maintenance*. Edited by Wallace R. Blischke,
690 D.N. Prabhakar Murthy; p. 423-446, ISBN: 0-471-41373-9, 2003, John Wiley and Sons,
691 Inc. <https://doi.org/10.1002/0471393002.ch19>
- 692 [34] Forys M.B., Kurowicka D., Peppelman B. (2013). A probabilistic model for a gas
693 explosion due to leakages in the grey cast iron gas mains *Reliability Engineering &*
694 *System Safety* volume 119, issue, year 2013, pp. 270 - 279.
695 <https://doi.org/10.1016/j.ress.2013.06.034>
- 696 [35] Brown A.J., Aspinall W.P. (2004). Use of expert opinion elicitation to quantify the
697 internal erosion process in dams. In Proc: *The 13th Biennial British Dams Society*

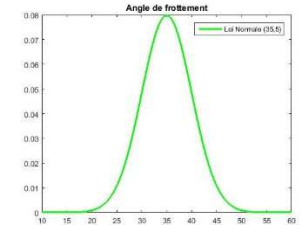
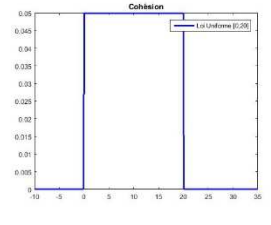
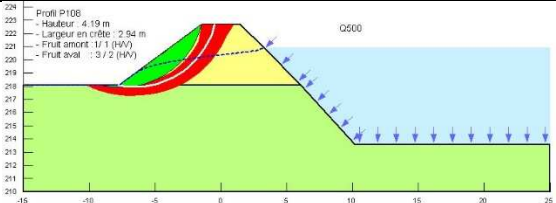
- 698 *Conference: University of Kent, Canterbury, 22-26th June 2004; 16pp*
699 (<http://www.britishdams.org/2004conf/synopses/brown.pdf>).
- 700 [36] Cooke R.M., Slijkhuis K.A. (2003). Expert Judgment in the Uncertainty Analysis of
701 Dike Ring Failure Frequency Case Studies in *Reliability and Maintenance*. Edited by
702 Wallace R. Blischke, D.N. Prabhakar Murthy; p. 331-352, ISBN: 0-471-41373-9, 2003,
703 John Wiley and Sons, Inc. <https://doi.org/10.1002/0471393002.ch15>
- 704 [37] Armstrong SJ, ed. (2001). *Principles of Forecasting: A Handbook for Researchers and*
705 *Practitioners* Kluwer Academic, Norwell, MA, 417–439.
706 <https://www.springer.com/us/book/9780792379300>
- 707 [38] Larrick R.P., Soll J.B. (2006). Intuitions about combining opinions: Misappreciation of
708 the averaging principle. *Management Science*. 52:111–127.
709 <https://doi.org/10.1287/mnsc.1050.0459>
- 710 [39] Lichtendahl KC Jr., Grushka-Cockayne Y., Winkler RL. (2013). Is-it better to average
711 probabilities or quantiles? *Management Science*. 59:1594–1611.
712 <https://doi.org/10.1287/mnsc.1120.1667>
- 713 [40] Colson A.R., Cooke R.M. (2017). Cross validation for the classical model of structured
714 expert judgment. *Reliability Engineering and System Safety* 163 (2017) 109- 120.
715 <https://doi.org/10.1016/j.res.2017.02.003>
- 716 [41] Quigley J., Colson A., Aspinall W., Cooke R M. (2018) Elicitation in the classical
717 model. In: LC Dias; A Morton; J Quigley, editors *Elicitation: The science and art of*
718 *structuring judgement*. Cham: Springer International Publishing; p. 15-36.
719 https://link.springer.com/chapter/10.1007/978-3-319-65052-4_2
- 720 [42] Carvajal C., Peyras L., Bacconnet C., Bécue J-P. (2009). Probability Modelling of Shear
721 Strength Parameters of RCC Gravity Dams for Reliability Analysis of Structural Safety.
722 *European Journal of Environmental and Civil Engineering*, Vol. 13, no.1. pp. 91-119.
723 <https://doi.org/10.1080/19648189.2009.9693087>
- 724 [43] Field, A. P. (2000). *Discovering statistics using SPSS for Windows: Advanced techniques*
725 *for the beginner*. London: Sage. <https://dl.acm.org/citation.cfm?id=518130>
- 726 [44] Vuillet M., Peyras L., Serre D., Diab Y. (2012). Decision making method for assessing
727 performance of large levee alignment *Journal of Decision Systems*, volume 21, no.2, pp.
728 137-160. <https://doi.org/10.1080/12460125.2012.680354>
- 729 [45] Lin, S-W., Bier V M. (2008). A study of expert overconfidence. 52:111–127. *Reliability*
730 *Engineering & System Safety* 93 711-721 <https://doi.org/10.1016/j.res.2007.03.014>
- 731 [46] Clemen, R.T., Lichtendahl, K.C. (2002). Debiasing expert overconfidence: a Bayesian
732 calibration model.PSAM6, San Juan, Puerto Rico.

733

734 **7 Appendix A: Variables of interest used in the case study**

735

736 The following four figures present an example questionnaire sheet for each failure mode
737 considered as a variable of interest.

Form 14		Cross-section – P 108	
<p>Given all the information available, what is the probability P_{fc} that the levee will fail due to sliding of the downstream slope in the case of a flood 500 (annual $P_{flood} = 1/500 = 2 * 10^{-3}$)? $P_{Total} = P_{fc} * P_{flood}$</p>			
Geometric data <i>Cf. Figure below</i>	<ul style="list-style-type: none"> - Height = 4.61 m - Altitude of crest = 222.69 m NGF - Width of crest = 2.94 m - Upstream batter = 1/1 (H/V) - Downstream batter = 3/2 (H/V) 		
Hydraulic data	<ul style="list-style-type: none"> - Flood level Q500 = 220.91 m NGF (1.78 m under the crest) - Hydraulic gradient level calculated in steady state 		
Geotechnical data	<p>Body and foundation</p> <ul style="list-style-type: none"> - Body : Sandy-gravel - Permeability $K = 10^{-3} m/s$ - Foundation Sandy-gravel, lenses of more or less compact materials. - Permeability $K = 10^{-6} \text{ à } 10^{-5} m/s$ - Unit weight: $\gamma_h = 21 kN/m^3$ 		
Probability laws of input data			
	<p>Angle of friction (body - foundation)</p> <p>Probabilistic law : normal law Mean : 35° Standard deviation : 5°</p>		<p>Cohesion (body - foundation)</p> <p>Probabilistic law : loi uniforme Min = 0 kPa Max = 20 kPa</p>
Flood : Q500			
Deterministic analysis	<p>SF Calculated (with mean values) Downstream slope 2,01</p>		
<u>Demand for response</u>			
Probabilistic analysis	Probability (P_{fc}) of sliding failure	P_{fc} : quantile q5%	
		P_{fc} : quantile q95%	
		P_{fc} : median value (q50%)	

738 **Figure 7: example questionnaire sheet for sliding**

739

740


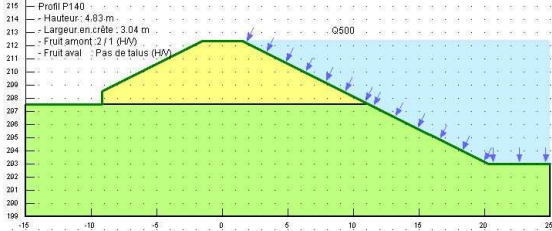
Form 17		Cross-section P 106	
<p>Given all the information available, what is the probability P_{fc} that the levee will fail due to internal erosion in the case of a flood 500 (annual $P_{flood} = 1/500 = 2 * 10^{-3}$)?</p> $P_{Total} = P_{fc} * P_{flood}$			
Aerial view of the section			
Cross-section			
Evaluation criteria presenting technical data on the entire section			
$C_{1,IE}$ Levee body permeability	<ul style="list-style-type: none"> - Body permeability : $K = 10^{-3} \text{ m/s}$ - Foundation permeability : $K = 10^{-6} \text{ to } 10^{-5} \text{ m/s}$ - Geometric Data : Height 3,24 m – Width of crest 4,61 m – Upstream batter 3H/2V – Downstream batter 1H/1V – flood level is 1.78 m under the crest for Q500 - Water leak : undetected - Sealing of the upstream protection: rip rap of the 50s 		
$C_{2,IE}$ Levee body resistance to internal erosion	<ul style="list-style-type: none"> - Materials constituting the levee body: sandy-gravelly, seem locally not very compact. - Materials constituting the foundation of the levee: sand-gravel, seem to present lenses of more or less compact materials. - Cohesion $c = 0 \text{ kPa}$ - Angle of friction $\phi = 35^\circ$ - Discontinuities of the embankment: not identified by the observation of an in situ geotechnician - The critical hydraulic gradient of materials is $i_{cr} = 0,1$ - The overall hydraulic gradient of this profile (P 106) is of $i_{glob} = 0.09$ (steady state) - The duration of floods 2 days 		
$C_{3,IE}$ Irregularities in levee body	<ul style="list-style-type: none"> - Trees and roots: significant presence of trees along the entire length of the dyke on the embankment side of the river. - Terriers: no significant burrows detected. - Crossing structures: no crossing works identified 		
Probabilistic analysis	Probability (P_{fc}) of internal erosion failure	P_{fc} : quantile q5%	
		P_{fc} : quantile q95%	
		P_{fc} : median value (q50%)	

741

Figure 8: example questionnaire sheet for internal erosion

742

Form 22		Cross-section – P 140	
<p>Given all the information available, what is the probability P_{fc} that the levee will fail due to overtopping in the case of a flood 500 (annual $P_{flood} = 1/500 = 2 * 10^{-3}$)?</p> $P_{Total} = P_{fc} * P_{flood}$			

<p>Aerial view of the section</p>								
<p>Cross-section</p>								
<p>Evaluation criteria presenting technical data on the entire section</p>								
<p>C_{1,0} Levee crest elevation</p>	<ul style="list-style-type: none"> - The embankment height is 4.83 m - The water is above the crest of 5 cm for a Q500 flood 							
<p>C_{2,0} Flow obstruction factors</p>	<ul style="list-style-type: none"> - No obstruction factors identified 							
<p>C_{3,0} Presence of low elevation points in crest</p>	<ul style="list-style-type: none"> - No low elevation point in crest identified 							
<p>C_{4,0} Operation of spillway(s)</p>	<ul style="list-style-type: none"> - No spillway of security along section 							
<p>C_{5,0} & C_{6,0} Resistance to overtopping of crest protection, landslide levee slope protection and waterside levee toe & Levee body resistance to overtopping</p>	<ul style="list-style-type: none"> - The duration of flood 2 days with overtopping for 3 hours - Height of downstream slope 3.04 m - Materials constituting the body of the dam: sand-gravel, seem locally not very compact. - Materials constituting the foundation of the levee: sand-gravelly, seem to present lenses of more or less compact materials. Permeability. - Cohesion of the body and the foundation $c = 0 \text{ kPa}$ - Angle of friction of the levee body $\phi=35^\circ$ - Angle of friction of the levee foundation $\phi=32^\circ$ - Etanchéité de la protection aval : terre végétale - Geometric irregularities: unidentified - Protection of downstream protection: topsoil - Geometric irregularities: unidentified 							
<p>Probabilistic analysis</p>	<p>Probability (P_{fc}) of overtopping failure</p>	<table border="1"> <tr> <td data-bbox="628 1583 1158 1617"> P_{fc}: quantile q5% </td> <td data-bbox="1163 1583 1375 1617" style="background-color: #f9cb9c;"></td> </tr> <tr> <td data-bbox="628 1624 1158 1657"> P_{fc}: quantile q95% </td> <td data-bbox="1163 1624 1375 1657" style="background-color: #f9cb9c;"></td> </tr> <tr> <td data-bbox="628 1664 1158 1688"> P_{fc}: median value (q50%) </td> <td data-bbox="1163 1664 1375 1688" style="background-color: #f9cb9c;"></td> </tr> </table>	P_{fc} : quantile q5%		P_{fc} : quantile q95%		P_{fc} : median value (q50%)	
P_{fc} : quantile q5%								
P_{fc} : quantile q95%								
P_{fc} : median value (q50%)								

743

Figure 9: example questionnaire sheet for overtopping

744

Form 19	Cross-section P 106
<p>Given all the information available, what is the probability P_{fc} that the levee will fail due to Scouring in the case of a flood 500 (annual $P_{flood} = 1/500 = 2 * 10^{-3}$)?</p> <p style="text-align: center;">$P_{Total} = P_{fc} * P_{flood}$</p>	


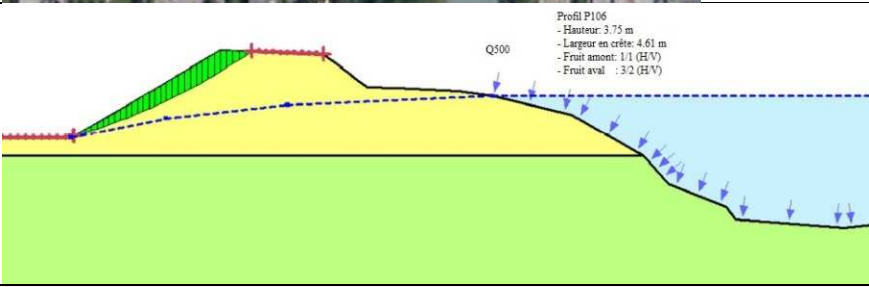
Aerial view of the section			
Cross-section			
Evaluation criteria presenting technical data on the entire section			
$C_{1,sc}$ & $C_{2,sc}$ Levee body protection resistance to external erosion on river side & Levee body resistance to external erosion	<ul style="list-style-type: none"> - Sealing of the upstream protection: pearl of the 50s - Height 3,75 m - Flood level Q500 1.25 m under the crest - The duration of floods 2 days - The average flow velocity of this profile (P 106) for a Q500 flood is 3.42 m / s - En générale, les vitesses d'écoulement dans le lit du Drac sont fortes, supérieures à 3,5 m/s pour le tronçon 1 - Body permeability : $K = 10^{-3} m/s$ - Materials constituting the levee body: sandy-gravelly, seem locally not very compact. - Cohesion $c = 0 kPa$ - Angle of friction $\phi = 35^\circ$ 		
$C_{3,sc}$ & $C_{4,sc}$ Foundation protection resistance to external erosion & Foundation resistance to external erosion	<ul style="list-style-type: none"> - Sealing of the upstream protection: pearl of the 50s - Permeability of the foundation : $K = 10^{-6}$ à $10^{-5} m/s$ - Materials constituting the foundation of the levee: sandy-gravelly, seem to present lenses of more or less compact materials. - Cohesion $c = 0 kPa$ - Angle of friction $\phi = 35^\circ$ 		
<u>Demand for response</u>			
Probabilistic analysis	Probability of scouring failure (P_{fc})	P_{fc} : quantile q5%	
		P_{fc} : quantile q95%	
		P_{fc} : median value (q50%)	

Figure 10: example questionnaire sheet for scouring

745

746

747