

Uncertainty and Expert Assessment for Supporting Evaluation of Levees Safety

Michel Hathout^{1,a}, Marc Vuillet^{1,b}, Laurent Peyras², Claudio Carvajal² and Youssef Diab¹

¹EIVP/Lab'Urba, 80 rue Rebeval, 75019 Paris, France

²Irstea, Research Unit RECOVER 3275 route Cézanne-CS 40064 F-13182 Aix-en-Provence Cedex 5, France

Abstract. In France, levees remain most of the time badly maintained; these long linear structures show signs of weaknesses on numerous occasions. Only incomplete information is usually available. The general lack of data describing the behavior of the infrastructure during unwanted events led to estimate their safety mainly from expert judgment. Thus the ability of the expert to predict the level of functioning of an infrastructure for a type of hazard and its intensity is crucial. An error of judgment can have very serious consequences and the production of reliable information requires the ability of the expert to report accurately the uncertainties in its estimations, as well as associated confidence. In order to meet this need, our research within Incertu project (French Ministry of Ecology funding) aims to produce relevant scientific approaches and tools for the collection and processing reliable experts' statements or combined with a confidence level in the context of uncertain information and input data.

1 Introduction

Levees of protection against flooding have a major role in the security of property and people. Unfortunately, disasters, those occur regularly in France and abroad, demonstrate whenever these structures are not infallible. The brutal characteristic of hazard aggravates the impact of flooding on people and other infrastructures.

In France, the park of levees remains unclear. The structures are poorly documented and data acquisition across the linearity is complex and expensive, due to the age and heterogeneity of most levees. Currently, the evaluation of the performances of these structures is made with expertise, based on knowledges about their structural condition and on hazard that could be affected. In practice [1], the engineer / expert evaluates qualitatively the performance of structures, usually without an explicit formalization of his reasoning. Considering challenges associated with these structures, French regulations, via risk studies, now request the probabilistic risk assessment of failure of these structures for several scenarios of flooding. Therefore the engineer must make a subjective assessment of failure probability of the structure, in the absence of complete physical model and absence of enough statistical series to be usable in terms of reliability.

However, the direct expression of the numeric data by the expert can sometimes lead to biased values [2, 3, 4]. These biases may be related to different factors. The judgments made by individuals based on heuristics [2],

that means informal reasoning, allowing the expert to quickly arrive at an evaluation or estimate, without the need of being really optimal. In a given situation, and to answer a specific question, the information by an expert can be a major source of information, but on the other hand it also leads to the necessity of developing approaches for the identification of bias, and if possible, to reduce them. Several approaches are described in literature concerning the familiarization with expert handling of probabilities in specific steps of elicitation or implementation of calibration procedures experts [5, 6].

Our scientific goal is to help identifying and reducing the biases that may affect the expert assessments of levees safety. We propose to answer to many scientific questions:

- What are the biases that can occur during the expert's evaluation of levees safety level?
- How to apply a calibration approach as proposed by Cooke and Goossens [6] in the context of levees, and what would these contributions be?

The studies of Vuillet [12] and Vuillet et al [1] propose some procedures to reduce bias that can occur when assessing the performance of levees. They are the starting point of our research in this project. Also, we'll present a brief summary of previous studies on the assessment of the safety level of operation of river levees and data imperfections typologies used to assess the safety of river levees. We'll also present a short summary of previous studies on the assessment of the safety level of operation of river levees and the typologies of

^a Corresponding author: michel.hathout@eivp-paris.fr

^b Corresponding author: marc.vuillet@eivp-paris.fr

imperfections of data used to assess the safety of river levees. Then, we'll present a summary of experts' bias and approaches described in the literature proposing to reduce them. Then, we'll present the application of the calibration model of Cooke [5] in the field of levees by interviewing experts on the permeability of various commonly constituent materials for embankments.

2 Methodologies for Assessment the Safety of Levees and Work of Vuillet et al

Several approaches exist for assessing the safety of levees. Fauchard and Mériaux [16] present a synthesis of classical steps constituting levees diagnosis: preliminary studies including analysis of the levees' history, topographical surveys, and detailed visual inspection. Geophysical and geotechnical knowledges are used to estimate the internal characteristics of the levees by studying variations of a physical field measured by longitudinal profiles. Numerical modeling is needed to calculate levees safety coefficients for certain failure mechanisms for profiles being the subject of substantial geotechnical knowledges. The qualitative assessment of levees by homogenous sections comes true by cutting linear units of incorporation and homogeneous loading. These diagnostic methodologies used primarily qualitative and do not take into account uncertainties remained of used data, nor aware of their impact on the determined safety level.

The assessment data of the security level of levees are subjected to many imperfections: uncertainties of longitudinal and transverse representativeness of an incomplete survey in the absence of observation levees in flood, etc. These imperfections, combined with the variability of the phenomena of levees degradation, prevent the vast majority of cases the engineer / expert to commit a precise estimate of the actual behavior of the structure subjected to more or less strong hydraulic solicitation. The term of imperfection generally describes the concepts of quality and quantity of available data. Typically there are three main types of data imperfections: uncertainty, vagueness and incompleteness [17].

In our context of levees, the uncertainty can be a random or an epistemic one [4]. Random uncertainties are related to natural variability of material properties, epistemic ones are related to the dissemination of investigations along linear and representativeness of the available data [1]. Inaccuracies are related to the measures of physical phenomena, conditions and format for the visual inspection of the structures. The inaccuracies of investigative materials or assumptions are based on geotechnical testing [1]. Incompleteness instead, is related to the fact that levees are still generally poorly known structures. The lack of construction plan, the rarity of critical flood and lack of work monitoring, and the absence of the levee manager recognizing lead in general to deficit of information.

Vuillet et al [1] propose a model for evaluating the performance of levees taking into consideration the

uncertainties in probabilistic format, it providing support to engineers for the diagnosis of structures for the main known failure scenarios. This model is intended to be used by a specialist engineer in rapid diagnostic status, by valuing all available data. The model provides support to the engineer for: identifying homogeneous portions of a linear of levees, to evaluate the performance of structures for different failure mechanisms and to specify the level of uncertainty of the produced results based on imperfect data available. Research process has three steps: (1) development of a functional model of the levees failure mechanisms, built from methods from Dependability and Qualitative Reasoning; (2) development of a support model deterministic decision including performance indicators for each failure mechanism levees, according to a method of construction of unique synthesis criteria; (3) development of a probabilistic performance assessment model including a method for taking into consideration the uncertainties of input information and results of the model in the context of subjective probabilities.

In this approach, the engineer intervenes to assess the implementation of the functions assigned to the components levees on a semi quantitative performance scale from 0 to 10. He gives his most likely evaluation together with a range of uncertainty, as a modal value and interquartile interval [5%; 95%], that allow to construct probability density functions that can be successively spread through Monte Carlo simulation until a level of performance of the segments considered accompanied by margins of uncertainty.

The approach also provides a path on processing through heuristics, mainly via a particular elicitation process and consulting a manual for handling subjective probabilities. In this state, these remain partial and still not possible to give some concrete elements to the biased nature of the valuations of an expert considering a particular area. Also, we present in the following sections the biases identified as likely to occur in the evaluation of the safety of levees and contribution of Cooke calibration model for quantified characterization of experts through a field of interest.

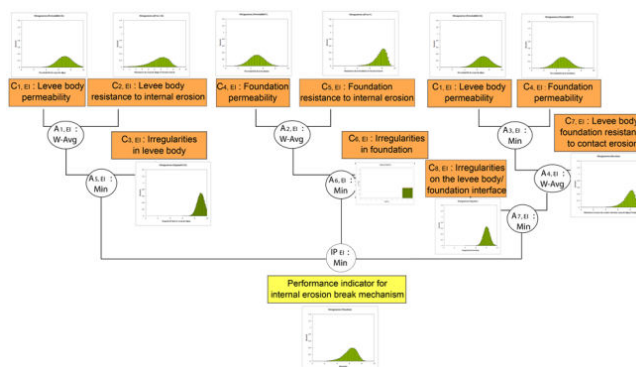


Figure 1. Uncertainty propagation criteria by Monte Carlo simulations for internal erosion performance indicator [1].

3 Judgment Experts and Bias

Many definitions exist to determine what an expert is and what differentiates him from the general population. For Ford and Wood [14], the expert is "A person capable of solving problems based on data from a specific case, on the general principles of the field, on strategies known in the field and on procedural knowledge or relationships". According to Meyer and Booker [7] the expert is "A person who has a strong knowledge of a particular field, he is recognized as qualified to answer questions by his peers or those conducting the analysis". In other words, an expert is a person with expertise and specific skills, and thereby having the ability to create, imagine and find links between different data (old and new) of risks that the return of experience alone does not permit; and this, in order to find answers and more relevant solutions to this risk, or at least anticipate an unknown behaviour during a risk.

An expert opinion or expert judgment is "the expression of opinion, based on knowledge and experience, the expert made it in response to questioning" [7]. More specifically, the judgment represents the state of knowledge of the expert during his answer at the question [8]. In general, we use the expert judgment in risk analysis, when data feedback is insufficient (in terms of quantity or quality) or nonexistent. They do not allow a statistical or probabilistic estimation precisely and / or consistently of considered elements [9]. In this case, a second source of knowledge, more subjective, can arise: the expert opinion or the expert judgment expressed by one or several experts (referred to an expert panel).

Despite their skills, depth knowledge and experience, experts are not immune to bias in their own opinions and probabilistic judgments [4, 10]. Indeed, there is always a transformation that can generate a difference between knowledge held and how it is rendered and expressed [11]. These differences are mainly expressed by the presence of bias that may occur at different times of the elicitation and analysis process, if precautions are not taken [2]. The elicitation corresponds to expert knowledge formalization process to enable them to save and / or share them [5].

The subjective evaluation of phenomena induced using heuristics. From a psychological point of view, the heuristics are "Rules of reasoning that lead to a simplification of the problem and solve it quickly but not always correctly" [13]. Thus, the biases are "Distortions between how we reason and that we should adopt to best ensure the validity of our inferences and conclusions" [13]. The Cognitive literature makes a relation between the heuristics and biases. It considers the heuristics like a generator of bias that may taint the skills of an expert to elicit his/her true opinion [2].

There are four main categories of bias [7]. (1) Cognitive bias (bias of the expert): due to intellectual limits and more generally to the tendency to simplify or make shortcuts when the problem is too complex. (2)

Motivational bias (bias condition): mainly induced by the hierarchical position of the expert or his expectations about the study. The expert does not carry its knowledge because of hierarchical pressures or the potential acceptability of answers. (3) Analyst bias: particularly involved in the choice of experts or in the interpretation of responses. (4) Technical bias: usually depend on the quality of questionnaires and tools used for collecting opinions.

Regarding cognitive biases, the literature of the field [2, 3] considers four main means: availability, performance, over-confidence and anchoring [1].

The availability bias is a mental shortcut considering only the most recent facts or overestimating their importance by reason of their "availability" in memory [4]. Availability bias in cases where one assimilates the engineer allows disrupting through one or more elements of "seers" than others. Example: the engineer will tend to give too much importance to a failure mechanism (among several possible mechanisms) at the origin of a recent disaster?

The representation bias is an error of estimation of the correspondence degree between a sample and a population [2]. This translates into a decrease in consideration of the conditional aspect of probability. Example: the engineer neglects the impact of ductwork on the safety of structures, while a significant percentage of levee failure is attributable to this type of phenomenon.

The overconfidence bias led the expert to minimize the uncertainty into his assessments. Conversely, pessimism leads the expert to give insufficiently informative distributions. Example: an overconfident engineer in the representativeness of a poll and minimizing the uncertainty associated on the available variability of materials.

The anchoring bias occurs when the expert leaves too much influence with a first spontaneous evaluation, followed by a difficult adjustment of its judgment. Example: after consultation of archived documents, the engineer evaluates a criterion according to the most likely mode regarding the first judgment and fails to adjust it regarding to the light of other elements contradicting this assessment.

4 Debiasing Approaches

Several approaches have been developed for the reduction or consideration of bias: a detailed presentation of the various bias for a better awareness of the expert [6], the use of specific elicitation procedures [3], the realization of training exercises [4, 6]. Regarding the evaluation of the safety of levees of protection against flooding, Vuillet et al [1] propose a combination of several of these approaches, mainly through an elicitation procedure and instruction manual of subjective probabilities.

Other treatment approaches of biases unexplored in the field of evaluating the levees safety propose the

calibration of experts' assessments from the experts' interrogations on the variables whose we know the results (calibration variables). The calibration models consist on assessing the consistency between the information produced by the expert and the known values (experimental /observed). Moreover, there is the consistency between values elicited by the expert and known values, and the score of calibration [5]. There are several models of calibration: the classic model of Cooke named the model "Cooke", the Bayesian model of Mendel-Sheridan named the model "MS", the determined weight model named the "Weights" model. Here we propose to detail the classical model of Cooke, which is probably the most commonly used model [6]. It can identify / treat cognitive biases holistically.

The classic model of Cooke [5] allows a mathematical linear combination of expert estimates provided in form of intervals. The final mathematical combination is based on a calibration step which aims to assess the quality and quantity of probabilistic assessments of experts. The calibration is based on the evaluation of the consistency between values elicited by the expert and known values. This results in determining an order among the experts by developing a system of rating (calibration scores) that will assign a weight to expert estimates from known cases.

The calibration score contains two main components: the component of calibration and the component of entropy. The calibration component C_e measures the quality of information given by the expert. It is based on measuring distances between a theoretical probability vector [0,1] and a calibration vector that is determined from a subjective probability vector (determined by the expert). This component gives a higher calibration score if the known value of the variable of calibration is into the interval elicited by the expert. The entropy component K_e measures the quantity of information given by an expert, in other words, his ability to be precise into his probabilistic estimations. It is based on measurement of distances between a subjective probability vector (given by the expert) and a theoretical probability vector on which expert probabilistic estimation was made. This component gives a lower score of calibration of the interval of uncertainty elicited by the expert is too broad. The calculation of these two components is based on notion outcome of "classical statistical": "Related information" between two vectors $I_e(c_i, p_i)$, this reflects the "classic" designation in the name of this model.

$$C_e = 1 - X^2(2 \cdot n \cdot I_e(c_i, p_i)) \quad (1)$$

$$K_e = \frac{1}{n} \sum_1^z I_e(p_i, s_i) \quad (2)$$

Where : s_i vector of subjective probabilities, p_i vector of theoretical probability, c_i vector of calibration, e number designating the expert, z number of inter-quartile spaces, $i \in [1, z]$, n numbers of calibration variables, $X^2(2 \cdot n \cdot I_e(c_i, p_i))$ distribution function of a variable $(2 \cdot n \cdot I_e(c_i, p_i))$

according to the Chi-square law, $I_e(c_i, p_i)$ information relating between a theoretical probability vector p_i and calibration vector c_i , $I_e(p_i, s_i)$ related information between the vector of subject probability s_i and the vector of theoretical probability p_i .

5 Application of the Classical Model of Cooke on the Levees

In this application, we implement the classic calibration model of Cooke. We propose to use as calibration variables the theoretical coefficient of permeability of materials consecutive to the levees body. By the calibration phase, characterize expert's opinion when asked to assess the permeability of a levee body (determining information for the evaluation of the safety of the structures face several failure mechanisms) [12]. Suppose an embankment levee consisted of two homogeneous sections respectively in *sandy-clay* and *sandy-loam* material. The most likely permeability value of a *sandy-clay* material corresponds to (10^{-8} m/s) and (10^{-6} m/s) for a *sandy-loam* material [15]. The expert panel contains seven experts: 1) - senior geotechnical engineer, 2)- senior engineer in river hydraulics, 3)- junior civil engineer, 4)- junior water engineer, 5)- junior geotechnical engineer, 6)- junior engineer in river hydraulics and 7)- The manager of levee. The objective is that each expert estimates the values of calibration variables on a scale compatible with coefficient of permeability ranging from (10^{-12} m/s) (Almost waterproof material) until 1 (High permeability materials) figure (2):

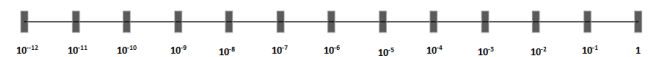


Figure 2. Scale of permeability coefficient (m/s).

The experts should give their estimations of inter-quartile range of values they consider credible [10%, 90%], and the value that they consider most likely 50%. Figure (3) below illustrates the subjective probability distributions of experts and their uncertainty intervals for the first calibration variable: the coefficient of permeability of *sandy-clay* materials. It also contains the known true value of the calibration variable (Real). We are limited to present the elicitations for the first calibration variable, since the elicitations for the second calibration variable are substantially the same as those of Figure 3.

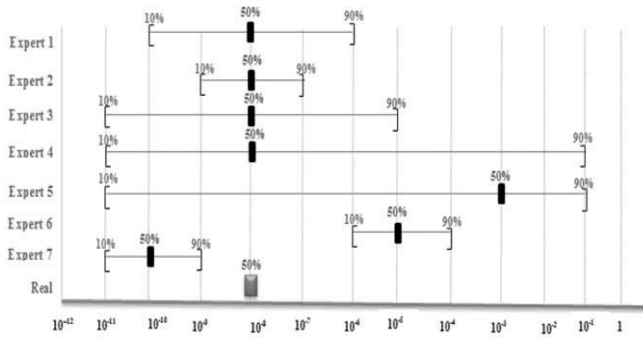


Figure 3. The subjective probability distributions of experts and their uncertainty intervals for the first calibration variable: the coefficient of permeability of sandy-clay materials.

The goal is to characterize the expert opinion for this case. We will then proceed to the calibration of subjective probability distributions elicited by the experts. At first, for calculating the calibration component C_e . We construct the calibration vector c_i for the elicitations of expert for each calibration variable. We calculate for each expert the information parameter related $I_e(c_i, p_i)$ of the vector of theoretical probability relative to the calibration vector. We talk about theoretical probability vector for describing the inter-quantile spaces [0%; 10%; 50%; 90%; 100%]. When using a large number of calibration variables, an expert opinion, theoretically, well-calibrated if the true value of the calibration variable is within in range of uncertainty elicited [10% ; 90%] in 80% of cases and 20% outside divided equally between the intervals [0 ; 10%] et [90% ; 100%]. At second, we calculate the entropy score of each expert by the entropy component K_e by calculating, for each expert, the relative information $I_e(p_i, s_i)$ of subjective probability vector s_i relative to theoretical probability vector p_i . Finally, with the latter two components a weight w'_e will be assigned to each expert translating that one of them has given the best evaluation and most accurate estimate. A relative weight w_e may also be defined based on the weight of all the experts. Table (1) presents the final results of this calibration phase.

	C_e	K_e	$w'_e = C_e \cdot K_e$	$w_e = w'_e / \sum w'_e$
Expert1	0.42	11.9	5.04	0.26
Expert2	0.42	12.7	5.36	0.28
Expert3	0.42	11.2	4.75	0.25
Expert4	0.42	8.5	3.59	0.19
Expert5	0.06	4.4	0.25	0.013
Expert6	0.004	6.5	0.03	0.001
Expert7	0.004	16.9	0.07	0.004

Table 1. The relative weight W_e assigned for each expert after a calibration procedure made according of classical model of Cooke (1991).

The results of the calibration phase show that experts 1,2,3,4, & 5 had a better score for component calibration than experts 6 & 7. We observe that the expert (2) had the best final calibration score (weight relative w_e , then the expert (1) and experts (3), (4) and (5). The experts (6) and (7) have the worst final calibration scores (relative weight w_e . In both cases, the true value of the calibration variable was outside of the ranges of uncertainty interval elicited by experts.

This calibration phase has allowed us to have via a score / weight an overall idea of bias of all experts' evaluation for a calibration data variable(s). Also these ratings / weights vary certainly more if we changed context of study. In addition, this model allows evaluating, via the component of calibration the quality of information probabilistic estimates of the expert, in other words, assess their accuracies. Via the entropy component, it also makes it possible to assess the amount of information contained in probabilistic estimates of the expert and therefore the expert's ability to be precise. We observe that the model assess the accuracy of the information given by the expert in respect of a single variable (here 10^{-8}), and not in relation to a value which can be described as strictly uncertain and an interval of uncertainty (including cases of natural variability). Thus, the model does not consider the uncertainty of the calibration variable or variable of interest, it does not therefore, as such, to describe the ability of the expert to model the uncertainty about the level of safety of levees. This is very important because, given the challenges faced and safety requirements related to these structures, it is usually better to take safety margins and to retain the most unfavorable / pessimistic assessment of the uncertainty range given by the expert on the safety of structures. We note that if we use a significant number of calibration variables, the results of experts should strive towards the theoretical probability vector to be well calibrated. Thus, via the components of calibration and entropy, the model allows a global description of the biased nature of the expert's assessments regarding to the calibration variable. Future research can allow refining this vision, for example, by integrating a possible range of uncertainty for the calibration variables.

6 Continuity and Application

This research, developed under the Incertu research project, 2013-2016(funded by MEEM, research program Risk Decision Territory 2013), associate EIVP, Irstea, the University Paris Est and as associate partners, the Department of Highways and Movement (DVD) of the City of Paris and the Isère Departmental Drac-Romanche -ID IDR Association); which manages nearly 200 km of levees of the Grenoble area and its surroundings. We plan to apply the presented approach for evaluating the probabilities of failure of real levees. The approach is to first assemble a panel of experts specialized in the evaluation of the safety of hydraulic structures, from Irstea Aix en Provence and the AD-IDR. The next step is to identify for our study the most interesting sections of

levees and synthesize information available in the form of a raised knowledge for the experts. In another step, the analyst propose scenarios of failure of sections built from methods of Dependability & Reliability, based on the research work of Serre et al. [18] and Vuillet et al [1], scenarios that will be, initially validated with the experts. The experts will be interviewed independently of each other anonymously, on calibration variables then, the probability of failure of the intermediate sequences of failure scenarios. Finally, evaluations obtained will be aggregated for each scenario and will be compared with the results of existing studies. The contribution of the method and its results can be well characterized and discussed concretely.

7 Conclusion and Perspectives

In risk analysis, when we have few data, it is common to rely on "expert judgment". This is particularly the case for the evaluation of the safety of levees to protect against flooding. Given the importantly risks associated with these structures, it appears necessary to have procedures to limit biases that may affect their assessments. We have identified that there are several kinds of bias and that it should combine approaches to identify and reduce them. We submitted a brief application of Cooke calibration model (1991) in the field of levees and showed the interest and the limitations of this model for our research approach. The model characterizes the trend for the adequacy and the precision of the experts for the calibration variable data. It does not allow to finely characterize the bias associated with the expression of expert uncertainties margins. Also, future research should be undertaken to refine the model.

8 References

1. Vuillet M., Peyras L., Carvajal C., Serre D., Diab Y., (2013). "Levees performance evaluation based on subjective probability". **European Journal of Environment and Civil Engineering**, Volume 17, issue 5/2013, pp 329-349.
2. Kahneman. D., Slovic, P., Tversky, A. (1982). «Judgement under Uncertainty, Heuristics and Biases». Cambridge University Press, Cambridge.
3. Morgan, M.G., Henrion, M. (1990). Uncertainty, a guide to dealing with uncertainty in quantitative risk and policy analysis, Cambridge University Press.
4. O'Hagan, A., Buck, C. E., Daneshkhan, A., Eiser, J. R., Garthwaite, P.H., Jenkinson, D.J., Oakley, J.E., Rakow, T. (2006). *Uncertains Judgments : Eliciting Experts' Probabilities*, West Sussex, UK, 338 p.
5. Cooke, R. (1991). «*Experts in uncertainty: opinion and subjective probability in science*». Oxford University Press, 336 p.
6. Cooke, R.M., Goosens, L. (2000). "Nuclear science and technology. Procedure guide for structured expert judgment". *Report to European commission. S. P. N. F. S. 1994-98*, Luxembourg Ruratom. Project report EUR 18820EN.
7. Meyer, M.A., Booker. J.M. (1993). «*Eliciting and Analyzing Expert Judgment: A practical Guide*», ASA-SIAM Series on Statistics and Applied Probability.
8. Ortiz, N.R., Wheeler, T.A., Breeding, R. J., Hora, S., Meyer, M.A., Keeney, R.L. (1991). "Use of Expert Judgment in NUREG-1150" *Nuclear Engineering and Design*, volume 126, pp. 313-331.
9. Muffat, S. (2005). «Professionnalisation du recueil des avis d'experts pour l'analyse des risques », Note EDF HT-52/05/023/A.
10. Pietrocattelli, S. (2008). «Analyse bayésienne et élicitation d'opinion d'experts en analyse de risques et particulièrement dans le cas de l'amiante chrysotile ». Mémoire de master, Université de Montréal, 110p.
11. Banton O. et Bangoy L. M. (1997) : Hydrogéologie : Multi-science environnementale des eaux souterraines. Université du Québec / AUPELF, 460p.
12. Vuillet, M. (2012). «Elaboration d'un modèle d'aide à la décision basé sur une approche probabiliste pour l'évaluation de la performance des digues fluviales », Thèse de doctorat, Université Paris Est, Ecole doctorale Ville, transport et territoires, 391 p.
13. Yachanin, S. A., Tweney, R. D. (1982). The effect of the thematic content on cognitive strategies in the four-card selection task. *Bulletin of the Psychometric Society*, 19, 87-90.
14. Wood, L. et Ford, J.M. (1993). « Structuring Interviews with Experts During Knowledge Elicitation », Ford et Bradshaw (dir.), Knowledge Acquisition as Modeling, New York, John Wiley and Sons Inc.
15. Castany G. (1982) : Principes et méthodes de l'hydrogéologie. Université de Pierre et Marie Crue (Paris VI), 233p.
16. Fauchard, C., Mériaux, P. (2004). Méthodes géophysiques et géotechniques pour le diagnostic des digues de protection contre les crues. Guide pour la mise en oeuvre et l'interprétation. Cemagref Editions, 124 p.
17. Bouchon-Meunier, B. (1995). La logique floue et ses applications. Addison-Wesley, 254 p.
18. Serre D., Peyras L., Tourment R., Diab Y., (2008), levee performance assessment: development of a GIS tool to support planning maintenance actions, **journal of infrastructure system**, asce, vol. 14, issue 3, pp. 201-213.