Fiabilité en fatigue des aubes des turbines hydrauliques dans un contexte imprécis

par

Mounia BERDAI

THÈSE PAR ARTICLES PRÉSENTÉE À L'ÉCOLE DE TECHNOLOGIE SUPÉRIEURE COMME EXIGENCE PARTIELLE À L'OBTENTION DE DOCTORAT EN GÉNIE
Ph. D.

MONTRÉAL, LE 06 SEPTEMBRE 2017

ÉCOLE DE TECHNOLOGIE SUPÉRIEURE
UNIVERSITÉ DU QUÉBEC

PRÉSENTATION DU JURY

CETTE THÈSE A ÉTÉ ÉVALUÉE

PAR UN JURY COMPOSÉ DE :

M. Souheil-Antoine TAHAN, directeur de thèse
Département de génie mécanique à l'École de technologie supérieure

M. Martin Gagnon, codirecteur de thèse
Institut de recherche d'Hydro-Québec (IREQ)

M. Mickaël GARDONI, président du jury
Département de génie de la production automatisée à l'École de technologie supérieure

M. Jean-Pierre KENNÉ, membre de jury
Département de génie mécanique à l'École de technologie supérieure

M. Daoud Aït-KADI, membre externe
Département de génie mécanique à l'Université Laval

M. François BEAUDOUIN, membre externe
Électricité de France (EDF)

ELLE A FAIT L'OBJET D'UNE SOUTENANCE DEVANT JURY ET PUBLIC

LE 31 AOUT 2017

À L'ÉCOLE DE TECHNOLOGIE SUPÉRIEURE

REMERCIEMENTS

J'aimerais remercier vivement toutes les personnes qui ont contribué de près ou de loin dans les différents travaux de cette thèse, particulièrement mon directeur de recherche, le professeur Souheil-Antoine TAHAN et mon codirecteur M. Martin GAGNON. J'aimerais leur transmettre ma profonde reconnaissance pour leur disponibilité, leur qualité d'encadrement et leurs précieux conseils.

J'adresse également mes très sincères remerciements au président et à tous les membres du jury d'avoir accepté d'évaluer ce travail. J'aimerais aussi remercier l'organisme MITACS-Canada et l'institut de recherche de Hydro Québec (IREQ) pour leur soutien et leur contribution financière, ce qui m'a permis de concentrer la majorité de mon temps à mes activités de recherche.

Mes remerciements les plus sincères sont surtout adressés à mes parents qui ont tout déployé pour ma réussite le long de toutes les étapes de ma vie. Aussi un gros merci pour ma petite famille pour m'avoir supporté durant cette période de recherche.

FIABILITÉ EN FATIGUE DES AUBES DES TURBINES HYDRAULIQUES DANS UN CONETXTE IMPRÉCIS

Mounia BERDAI

RÉSUMÉ

La fiabilité des turbines hydroélectriques est une fonction complexe qui dépend principalement des propriétés mécaniques des matériaux et des cycles de chargement. Ces propriétés intrinsèques aux matériaux sont affectées par le vieillissement et par les conditions d'exploitation. Par conséquent, leurs valeurs originales considérées lors de la conception des turbines hydrauliques ne peuvent être utilisées le long de la vie utile de ces équipements et d'où le recours aux avis des experts pour leurs mises à jour. Dans ces cas, les experts peuvent se baser sur les théories probabilistes ou sur les théories de probabilité imprécises, pour formuler leurs avis et mettre à jour ces propriétés mécaniques.

Dans l'article # 1, on analyse comment ces théories affectent le calcul de la fiabilité en fatigue, pour un modèle utilisant l'approche FORM (First Order Reliability Method) et ayant un état limite défini par le diagramme de Kitagawa-Takahashi. Dans cette contribution nous avons proposé une approche pour étendre le calcul de fiabilité sur des variables exprimées selon la théorie des probabilités imprécises. Nous avons aussi identifié que les données élicitées selon des distributions bornées, limitent la précision du modèle. Pour contourner cette limitation, une approche qui imite les distributions non-bornées et respectant le comportement physique des variables a été suggérée. L'article conclut que les théories de modélisation des avis des experts sont équivalentes et que la priorité devrait être accordée aux avis formulés selon des distributions non bornées.

Pour formuler leurs avis, les experts suivent généralement des techniques appropriées et accoutumés aux sujets étudiés. Les exemples des techniques d'élicitation proposées dans la littérature contrôlent et encadrent souvent les avis d'experts, en les orientant vers un consensus ou vers un choix spécifique. Autrement, en absence de ces encadrements les experts formulent leurs opinions selon leurs propres connaissances et selon leur compréhension du sujet, ce qui peut mener dans certains cas à des avis disjoints ou totalement opposés. Une situation qui sera d'autant plus compliquée lorsque les experts sont invités à prédire des données qui ne possèdent pas des valeurs de référence. Dans le cadre de l'article #3, nous avons exploré quelques catégories des techniques d'élicitation, avec et sans support. Les résultats obtenus montrent que pour les domaines où les données à éliciter ont un riche historique, une technique d'élicitation avec support sera recommandée dans le but de limiter la variation entre les avis des experts.

Finalement, pour les systèmes multi-variables, nous pouvons nous retrouver avec plusieurs experts pour l'élicitation de chacune des entrées du système. Dans cette situation, on se

demande quelle est la meilleure façon de combiner ces données : Faut-t-il les combiner avant leur propagation dans le modèle étudié ou bien les combiner après avoir propagé chaque opinion séparément dans le modèle? Dans l'article #2, nous avons exploré certains paramètres pouvant affecter la différence entre ces deux modes d'agrégation. Dans ce sens nous avons proposé une métrique δ pour la quantification de la divergence entre les opinions des experts. Nous avons également suggéré l'utilisation de la moyenne des fonctions de répartition comme règle d'agrégation. En effet cette moyenne semble appropriée pour la combinaison des opinions exprimées selon les distributions probabilistes ou exprimées selon les distributions non probabilistes. Aussi l'adoption de cette règle d'agrégation permet d'éviter les limitations rencontrées avec l'approche FORM, lorsque les opinions des experts sont exprimées selon des distributions bornées (article #1)(Berdai, Tahan et Gagnon, 2016). Les résultats de cette étude montrent que pour le modèle de fiabilité étudié, la différence entre les deux modes d'agrégation devient significative seulement lorsque le point d'exploitation est dans la région sûre, proche de l'état limite du modèle de propagation.

Mots-clés: Fiabilité des turbines hydroélectriques, élicitation des opinions des experts, approche des quantiles, distribution bornée, métrique de divergence, techniques d'élicitation, coefficient du levier

FATIGUE RELIABILITY OF HYDRAULIC TURBINE BLADES IN AN IMPRECISE CONTEXT

Mounia BERDAI

ABSTRACT

The reliability of hydroelectric turbines is a complex function which depends mainly on the mechanical properties of the material and loading stress. These properties are affected by aging and operating conditions; therefore their original values considered during the hydraulic turbines design cannot be used along the useful life of the equipment. Hence, the need to the use of expert opinions to update these proprieties. In such cases, experts may rely on probabilistic theories or imprecise probabilities to formulate their opinions and update these properties.

In paper # 1, we analyze how these theories affect the reliability calculation based on the FORM (first order reliability method) approach and having the Kitagawa-Takahashi diagram as a limit state. In this contribution we proposed an approach to extend the reliability calculation on variables expressed according to the imprecise probability theories. Also for the studied model, we highlighted that the variables expressed according to bounded distributions, reduce the model accuracy. To avoid this limitation, an approach that imitates unbounded distributions and respecting the physical behavior of the required variables has been suggested. The paper concludes that the modeling theories used to formulate expert opinions are equivalent and the priority should be given to opinions based on unbounded distributions.

In order to formulate their opinions, the experts generally follow some elicitation techniques, appropriate to the studied subject. Examples of elicitation techniques proposed in the literature often control expert opinions by guiding them towards a consensus or to a specific choice. Otherwise, in the absence of these frameworks, experts formulate their opinions according to their own knowledge and according to their understanding of the subject, which may lead in some cases to disjoint or totally different opinions. This situation will be more complicated if experts have to predict data without any reference values. In paper # 3, we proposed and compared some elicitation techniques and obtained results showed that for domains where the required variable has a rich history, elicitation techniques with supports will be recommended in order to limit the variation between expert opinions.

For multi-variable systems, we can have several experts available for the elicitation of each system input. In this situation, we wonder what the best way to combine these data is: before their propagation in the system model; or combining them after the propagation of each opinion separately in the system model. In paper # 2, we explored some parameters that can affect the difference between these two aggregation modes. In this sense we proposed the divergence metric δ for the quantification of the divergence between the expert opinions. We also suggested the use of the Cumulative Distribution Averaging as an aggregation rule and which seems appropriate for opinions expressed according to probabilistic or non-probabilistic distributions. This aggregation rule allows also avoiding the limitations encountered with the FORM approach (article#1) (Berdai, Tahan et Gagnon, 2016), when expert opinions are expressed according to bounded distributions. Findings from this study showed that for the studied reliability model, the difference between the two aggregation modes becomes significant only when the operating point is in the safe region, close to the model limit state. In this case, aggregation before propagation is more conservative than aggregation after propagation.

Keywords: Reliability of hydraulic turbines, expert opinions elicitation, quantile approach, bounded distribution, Divergence Metric, leverage coefficient.

TABLE DES MATIÈRES

Page

LISTE DES TABLEAUX

Page

LISTE DES FIGURES

LISTE DES ABRÉVIATIONS, SIGLES ET ACRONYMES

ASME	American Society of Mechanical Engineers
CDA	Cumulative Distribution Averaging (CDA)
CDF	Cumulative density function
DM	Divergence metric
EVT	Extreme value Theory
FORM	First Order Reliability Methods
GIS	Geographic information system
HCF	High Cycle Fatigue
KEEJAM	Knowledge Engineering Methodology for Expert Judgement Acquisition and Modelling
LCF	Low Cycle Fatigue
LDD	Linguistic descriptions of data
LEFM	Linear Elastic Fracture Mechanics
MPI	Magnetic Particule Inspection
MWh	Mega Watt heure
NDD	Natural language generation
NDT	Non Destructive Technique
NUREG	Nuclear Regulatory
POD	Probability of Detection
SHELF	SHeffield ELicitation Framework
SEPT	Standard for Educational and Psychological Testing

TBM	Transfer Belief model

LISTE DES SYMBOLES ET UNITÉS DE MESURE

a	Taille du défaut [mm]
a_{min}	Lower bound of a possibility distribution with a triangular shape (article#1)
a_{max}	Upper bound of a possibility distribution with a triangular shape (article#1)
α	Niveau toléré de l'erreur du type I
$\alpha - cut$	Processus d'échantillonnage de certaines formes des probabilités imprécises (exemple théorie des possibilités)
β	Represents the Gumbel scale factor
β_a	Reliability index after propagation
β_b	Reliability index before propagation
β_{HL}	Indice de fiabilité selon Hasofer-Lind
c	Possibility distribution mode (article#1)
D	The support domain (article#2)
d_{inf}	The lower bound of the support domain D
δ	Divergence Metric
d_{sup}	The upper bound of the support domain D
ΔK_{th}	Seuil de propagation des fissures
$\Delta\sigma_0$	Limite d'endurance [MPa]
$\Delta\sigma$	Contraintes de chargement [MPa]
$\Delta_{j,k}$	Interaction matrix
E_i	Expert i
φ	Fonction de répartition de la loi normale
F_{CDA}	Cumulative Distribution Averaging (CDA)

f_D	Probability distribution D characterised by a probability density function f_D
f_{D-e}	The extended distribution.
F^e	The cumulative density function of physical data
F_i	The cumulative distribution associated to the probability law P_i
F^m	The cumulative density function of estimated data
f_P	The probability density function of the probability law P
F_G	Cumulative Gumbel distribution
F_U	Cumulative Uniform distribution
F_T	Cumulative Triangular distribution
$g(x)$	État limite du modèle de fiabilité pour le vecteur $\boldsymbol{x}$
G0i	G01, G02, G03, G04 G05 points in the probabilistic space (article #1)
Gi	Operating points (G1, G2, G3 and G4)
K_t	Facteur de concentration des contraintes
l_i	Leverage coefficient
μ	The Gumbel localization factor
n	Number of experts (article #2)
N	Sampling levels
O_i	The opinion provided by the expert E_i
ω_i	Weight of expert E_i
$\boldsymbol{\Omega}$	Support domain common to all collected expert opinions (article#3)
$\boldsymbol{\Pi}_i$	Matrix with CDF of expert E_i opinions for each elicitation technique j and each operating condition k
$\boldsymbol{\pi}$	Variation matrix

P_f	Probabilité de défaillance
R	Ratio de chargement
σ_{LCF}	Loading stress corresponding to *Low Cycle Fatigue* [MPa]
σ_{HCF}	Loading stress corresponding to *High Cycle Fatigue* [MPa]
u_1	Standardized variable associated to the defect size (a)
u_2	Standardized variable associated to the loading stress ($\Delta\sigma$)
$x_{\frac{\alpha}{2}}$	Quantile corresponding to $\frac{\alpha}{2}$ level
$x_{1-\frac{\alpha}{2}}$	Quantile corresponding to $1-\frac{\alpha}{2}$ level

INTRODUCTION

Définition de la problématique

La déréglementation des marchés énergétiques et l'avènement des centrales éoliennes mènent à des fréquences élevées des démarrages/arrêts des turbines hydrauliques, ce qui accélère leurs fatigues et influence leurs plans d'exploitation et de maintenance. Or, pour une exploitation dite optimale, les gestionnaires doivent disposer d'informations précises reflétant, à un instant donné, la fiabilité résiduelle de ces turbines hydroélectriques. Ce calcul de fiabilité dépend de certains paramètres tels que : la nature du chargement, les limites du matériau, le niveau de dégradation et du dommage cumulé ainsi que de l'historique de maintenance et des réparations. Ces paramètres sont de nature aléatoire, de ce fait leurs effets sur le calcul de fiabilité seront assurément mieux décrits par un modèle probabiliste. Dans ce sens Gagnon *et al.* ont proposé un modèle pour le calcul de la fiabilité en fatigue des aubes des turbines hydroélectriques (Gagnon et al., 2013), basé sur l'approche de Hasofer-Lind. L'indice de fiabilité ainsi obtenu (β_{HL}), permet d'estimer la probabilité d'être dans la zone 'non-sûre' P_f et qui est calculée selon l'approximation linéaire du premier ordre FORM (*First Order Reliability Methods*) de la fonction décrivant l'état limite du modèle $g(x)$ (Figure 2.1). Cette probabilité est donnée par l'équation: $P_f \approx 1 - \varphi(\beta_{HL}) = \varphi(-\beta_{HL})$ où φ est la fonction de répartition de la distribution normale.

L'état limite $g(x)$ du modèle proposé par Gagnon *et al.* est défini par le diagramme de Kitagawa-Takahashi qui est caractérisé par deux seuils: le premier est le seuil de propagation des fissures $\Delta K_{th}[\mathrm{MPa}\sqrt{\mathrm{m}}]$ déterminé par la mécanique de rupture linéaire (LEFM) et le deuxième seuil correspond à la limite de fatigue $\Delta\sigma_0[\mathrm{MPa}]$ (Figure 2.1 (a)). L'espace sous ces deux seuils représente la zone sûre d'exploitation où il y a une propagation très faible de fissures ou pas de propagation (Gagnon et al., 2013).

Cet état limite $g(x)$ dépend des propriétés mécaniques des matériaux qui sont influencées par le vieillissement et par les conditions d'exploitation (Thibault, Gagnon et Godin, 2015). Par suite, les valeurs originales de ces propriétés, utilisées lors de la conception des turbines, ne peuvent pas être utilisées tout le long des vies utiles des turbines. Ainsi, il est essentiel de mettre à jour ces paramètres par de nouvelles mesures, provenant soit des opérations d'inspection ou bien issues de l'élicitation des avis des experts, pour améliorer l'estimation de la fiabilité.

Le calcul de l'indice de fiabilité selon le modèle proposé par Gagnon *et al.* nécessite aussi la connaissance de deux autres variables aléatoires : la taille du défaut dans les zones à hautes contraintes mécaniques et l'amplitude des contraintes de chargement. Cependant, l'accès à des valeurs *à jour* pour ces deux intrants est souvent très laborieux et parfois difficile à cause des contraintes d'exploitation des parcs hydrauliques. Dans ce cas, leur mise à jour peut s'effectuer selon des modèles physiques issus de la mécanique des milieux continus. Toutefois, l'élaboration de tels modèles peut s'avérer un processus lent et assez dispendieux; ce qui justifie souvent le recours aux avis des experts qui s'apprêtent comme une alternative peu coûteuse. Cette question d'incorporation des avis d'experts dans le processus d'estimation de fiabilité, selon le modèle proposé par Gagnon *et al.*, n'a pas été abordée auparavant et sera traitée dans le cadre de cette thèse.

Les avis fournis par les experts sur une même variable ne sont pas systématiquement identiques, puisque les experts n'adhèrent pas généralement aux mêmes écoles de pensées et ne possèdent pas exactement les mêmes connaissances et expériences. La diversification des théories de modélisation et leurs différences dans la représentation de l'information peuvent aussi être pointées comme facteurs expliquant cette différence dans les avis des experts. D'où notre motivation pour comparer certaines de ces théories afin d'évaluer leur impact sur le calcul de fiabilité.

Pour réduire ces écarts dans leurs avis, les experts suivent souvent des techniques d'élicitation pour formuler leurs opinions. Dans la littérature, on distingue deux catégories de techniques d’élicitation: des techniques directes et des techniques indirectes. Des comparaisons entre certaines techniques d’élicitation révèlent que sur un même sujet les opinions obtenues de l’élicitation peuvent différer significativement, pendant que d’autres recherches stipulent que les techniques d’élicitation peuvent être équivalentes sous certaines conditions. Face à ce dilemme, il est difficile de se prononcer sur l’équivalence (ou non) des techniques d'élicitation, surtout lorsque nous changeons de domaine d’étude ou des critères d'évaluation. D'où l'idée d’analyser certaines techniques pour évaluer leurs reproductibilités et identifier, si possible, celles qui s’adaptent le mieux à la prédiction de la taille des défauts, qui est la principale entrée du modèle de fiabilité de Gagnon *et al.* (Gagnon et al., 2013). La discussion des différentes catégories des techniques d’élicitation est effectuée dans notre revue de l’état d’art au CHAPITRE 1 et dans le CHAPITRE 4 (article #3).

Finalement, pour certains systèmes multi-variables, nous pouvons nous retrouver avec plusieurs experts pour éliciter chacune des entrées du système. Dans cette situation, on s’interroge sur la meilleure façon de combiner les avis ainsi collectés; faut-t-il les combiner avant leur propagation dans le modèle du système, ou bien les combiner après avoir propagé chaque combinaison d’opinions séparément dans le modèle de propagation du système?
La différence entre les deux modes d'agrégation (avant propagation et après propagation) ainsi que leurs effets sur les calculs de fiabilité sont méconnus, précisément lorsque le modèle de propagation est non linéaire.

Ces questions et d’autres constituent le cœur de notre projet de recherche.

Objectifs du projet

Typiquement les avis des experts sont sollicités quand il y a un manque de données ou bien quand on désire améliorer la prise de décision. Dans le cas du modèle de fiabilité basé sur l’état limite du diagramme de Kitagawa-Takahashi, l’élicitation de l’avis des experts devient une opération incontournable pour mettre à jour les paramètres de ce modèle. Dans ce cas, les avis des experts peuvent être exprimés selon des théories probabilistes ou encore selon les théories des probabilités imprécises. Les récentes recherches fournissent des résultats intéressants concernant l'utilisation des approches issues de la famille des probabilités imprécises, mais ne montrent pas comment ces théories affectent le calcul de probabilité ni comment leur interaction avec les autres théories de probabilité classiques influence le calcul de fiabilité. Le premier objectif de cette recherche est de comparer certaines de ces théories, notamment: la théorie des intervalles, la théorie de possibilité et la théorie de l'évidence. Puis d'évaluer leurs effets sur le calcul de fiabilité basé sur l'approche de Hasofer-Lind et ayant comme état limite le diagramme de Kitagwa-Takahashi.

D'un autre côté, pour cerner la dispersion qui peut avoir lieu entre les avis des experts, causée par le manque d'encadrement, on propose de comparer certaines techniques d'élicitation en vue de mettre en évidence leur reproductibilité et les variations entre elles et, si possible, d'identifier celles qui semblent plus appropriées au modèle de fiabilité proposé.

Ainsi, les objectifs de ce projet visent à répondre aux questions suivantes (Figure 0.1):

- Comment les théories de modélisation de l'avis des experts selon les théories probabilistes et les théories de probabilité imprécise affectent le calcul de la fiabilité en fatigue ? Quel est l'effet de l'utilisation d'un 'mélange' de ces théories sur le calcul de fiabilité lorsque le système est décrit par un modèle multi-variables avec un état limite non explicite ? Ces questions sont traitées dans le CHAPITRE 2 qui représente l'article #1 et qui a été publié dans *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems*.
- Dans le cadre de cette recherche on a examiné aussi la différence entre la combinaison des avis des experts avant la propagation dans le modèle du système et leur combinaison après la propagation dans le modèle du système. Cette question est traitée au CHAPITRE 3 qui représente l'article #2 et qui a été soumis au *Journal Information Sciences.*
- Finalement, est-ce que les techniques d'élicitation sont reproductibles et comment elles influencent les avis des experts? Comment aider les experts à exprimer leurs connaissances sans les biaiser ? Ces questions sont traitées dans le CHAPITRE 4 qui représente l'article #3 et qui a été soumis au *Journal of Risk Research.*

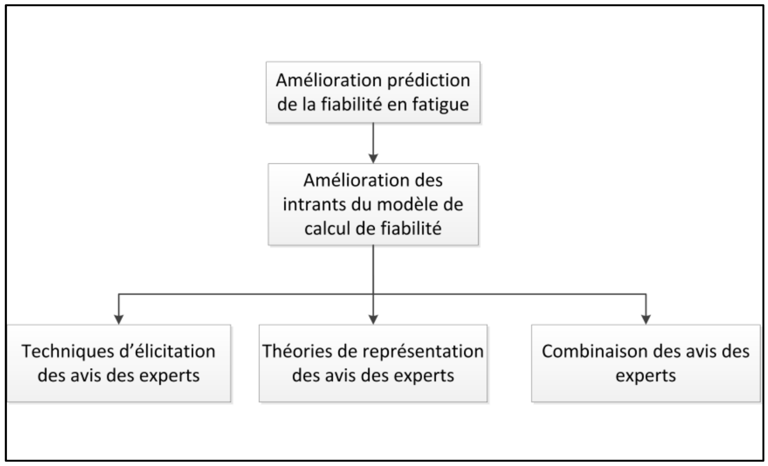


Figure 0.1 Structure de la problématique de recherche

Structure de la thèse

Cette thèse est organisée par articles, chaque article est publié ou soumis à un journal avec comité de lecture. Chacun des articles est inclut dans un chapitre distinct et répond à un des points précisés dans la section précédente:

- INTRODUCTION: Introduction à la problématique de recherche.
- CHAPITRE 1: Revue de littérature.
- CHAPITRE 2: Probabilités imprécises dans l'évaluation de la fiabilité en fatigue.
- CHAPITRE 3: Comparaison des techniques d'agrégation avant propagation et agrégation après propagation.
- CHAPITRE 4: Techniques d'élicitation appropriées pour le modèle de fiabilité en fatigue.

Suite à l'introduction, une revue de la littérature est effectuée afin de mettre l'emphase sur l'état de l'art et sur les travaux récents effectués par la communauté scientifique, pour chacun des domaines visés dans la section des objectifs. Ensuite, dans le CHAPITRE 2, nous proposons une approche pour la comparaison des théories issues des probabilités imprécises, fondée sur un ensemble de cas d'étude. Nous avons proposé également une approche pour étendre le calcul de fiabilité selon la méthode Hasofer-Lind et le diagramme de Kitagawa-

Takahashi, sur des variables dérivées de probabilités imprécises. Les résultats obtenus montrent qu'il existe une certaine homogénéité entre les théories considérées. L'étude conclut aussi que dans le cas circonstancié du calcul de fiabilité selon Hasofer-Lind et le diagramme de Kitagawa-Takahashi, la priorité devrait être accordée aux opinions des experts exprimés en termes de distributions non-bornées.

Au CHAPITRE 3, on étudie l'aspect relié au scénario de fusion des avis des experts: fusionner les avis comme une seule variable à l'entrée puis la propager, ou bien effectuer la propagation des différents avis et ensuite fusionner les sortants du modèle de propagation? La situation sera plus compliquée quand le système a plus d'une variable d'entrée, car dans ce cas on devra considérer toutes les combinaisons possibles de tous les avis d'experts, associés à chacune des variables d'entrées du système. Dans ce sens, les résultats obtenus de notre recherche, basée sur le modèle de fiabilité de Gagnon *et al.* (Gagnon et al., 2013), montrent que dans la zone sûre proche de l'état limite du modèle de propagation, l'agrégation avant propagation est généralement plus conservatrice que l'agrégation après propagation.

Le CHAPITRE 4 résulte d'un corollaire logique de notre revue de la littérature où nous avons constaté qu'il est difficile de décider si les techniques d'élicitation sont interchangeables ou non, surtout lorsqu'on change le domaine d'étude ou on change les critères d'évaluation. Ainsi, nous discutons un certain nombre de techniques afin de les analyser et de plébisciter, si possible, la technique d'élicitation appropriée pour l'élicitation des paramètres d'entrées du modèle de fiabilité basé sur le diagramme de Kitagawa-Takahashi. Dans le même objectif, nous avons proposé 4 techniques d'élicitation que nous avons comparées. Les résultats obtenus de cette comparaison ont montré que la technique C et la technique D, basées sur une représentation probabiliste, semblent limiter la variation entre les avis des experts, dans les cas du modèle étudié. Rigoureusement, nous ne pouvons pas affirmer qu'elles sont forcément « les meilleures » car dans le cas de prédiction (sans valeurs de références) elles risquent de biaiser les avis des experts. De ce fait, il nous semble qu'il est primordial de considérer des critères additionnels pour l'évaluation des techniques d'élicitation.

Finalement, dans la Conclusion, nous résumons l'apport de chacun des articles pour l'amélioration des méthodes d'estimation de la fiabilité en fatigue des turbines hydroélectriques et nous présentons par la suite, les perspectives pour de futures recherches.

CHAPITRE 1

REVUE DE LITTÉRATURE

1.1 Avant-propos

Le projet de recherche actuel vise à améliorer la gestion des parcs hydrauliques en se basant sur l'indice de fiabilité de Hasofer-Lind, déterminé à l'aide du modèle proposé par Gagnon *et al.* (Gagnon et al., 2013). Dans ce sens, dans la partie Introduction on a identifié certaines pistes menant à l'amélioration de la qualité de l'estimation de la fiabilité en fatigue. Ce qui permet par conséquent de rencontrer notre objectif ultime; celui de l'optimisation de l'exploitation des parcs hydrauliques. Ces pistes d'amélioration se rapportent à différents domaines de recherche ce qui reflète la nature multidisciplinaire du projet dans lequel on fait appel, entre autres, aux : théories de modélisation et de propagation des incertitudes épistémiques, processus et techniques d'élicitation des avis des experts ainsi qu'au domaine de la fusion des avis des experts (Figure 1.1). Par conséquent, notre revue de l'état d'art sera concentrée sur les domaines de fatigue, de modélisation des incertitudes et des processus de l'élicitation des avis des experts. Dans ce qui suit, on présente un portrait succinct de chacun de ces domaines. Un 'état de l'art' plus développé est introduit dans chacune de nos trois contributions (publications).

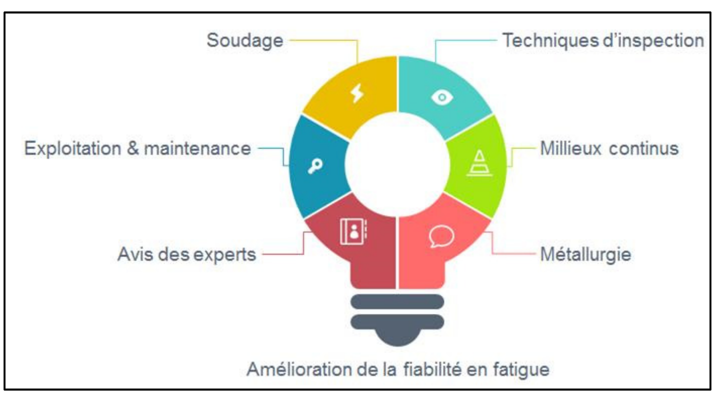


Figure 1.1 Disciplines impliquées dans le calcul de fiabilité en fatigue

1.2 Modélisation de la fatigue dans les turbines hydrauliques

La maintenance des roues des turbines hydrauliques consomme environ 17% du nombre total d’heures réservées à l'entretien complet des turbines (Aoudjit, 2010). Ce nombre élevé des heures de maintenance s'explique principalement par les coûts de réparation dus aux problèmes de fatigue et de cavitation.

En effet des recherches ont indiqué que dépendamment du choix des points de fonctionnement (hauteur de la chute, ouverture des vannes guides, débit) et de la procédure de démarrage des turbines hydrauliques, on peut réduire ou accélérer significativement le processus de fissuration et de fatigue des aubes (Allan et Roman, 1989; Bourdon et al., 1999). Ainsi, un choix judicieux des séquences de démarrage et des points de fonctionnement pourraient réduire les fluctuations de pression et du débit, qui sont les principales sources des vibrations (et des contraintes dynamiques) dans les structures de la turbine. En fait, certains points de fonctionnement génèrent plus de pression et de débit que d'autres, ce qui conduit à la génération de cycles dites HCF « *High Cycles Fatigue* »; que différentes études pointent comme étant le principal mécanisme d’initiation et de propagation des fissures (Brekke, 2013; Frunzăverde *et al*., 2010; Huth, 2005; Xiao, Wang et Luo, 2008). La forme de l’aube et la zone de transition (entre l’aube et la bande ou la couronne) ont aussi un effet sur la distribution des contraintes et sur le facteur de concentration des contraintes K_t (Bergmann-Paulsen, 2012; Huth, 2005; Thapa et al., 2012). Toutefois, le profil de l’aube est un paramètre prédéfini lors de la conception de la roue et que l’exploitant ne peut pas contrôler, contrairement aux séquences d’ouverture et aux points de fonctionnement.

En plus des contraintes de chargement (régimes permanent et transitoire), les caractéristiques mécaniques de la roue de la turbine hydraulique varient aussi en fonction de l’historique de fabrication et des réparations qui sont le principal responsable des contraintes thermiques (résiduelles) dans les zones soudées. En effet, les cycles thermiques résultant du soudage et du traitement thermique, provoquent un échauffement local non uniforme, par conséquent une déformation hétérogène dans le matériau en question et une augmentation de la sensibilité de la zone soudée à la fissuration. Les distributions de ces contraintes résiduelles dépendent en partie : des propriétés des matériaux, des procédures de soudage incluant les conditions de retenue, l'apport de chaleur et le nombre de passes de soudage (Brickstad et Josefson, 1998; Deng et Murakawa, 2006). Thibault *et al.* avancent que le traitement thermique post-soudage peut être prometteur pour baisser ces tensions résiduelles (Thibault, Bocher et Thomas, 2009). Aussi, ils préconisent d'effectuer des inspections ultrasons après chaque réparation pour détecter les défauts engendrés par le soudage (Thibault, Bocher et Thomas, 2009). En absence de ces mesures, les contraintes thermiques fragilisent localement le matériau et affectent ses propriétés mécaniques. C'est le cas du seuil de propagation de la fissure $\Delta K_{th}[\mathrm{MPa}\sqrt{\mathrm{m}}]$ qui conditionne la propagation des fissures et qui à son tour est affecté par le ratio du chargement R, la température d'exploitation, la taille du grain, le niveau d'oxydation et l'environnement corrosif (voir Figure 1.2) (Richard W. Hertzberg, 2013; Taylor, 1988). Par exemple, en présence d'environnement corrosif et un ratio de chargement R assez bas, les couches d'oxyde peuvent s'accumuler entre les faces opposées de la fissure et mener à sa fermeture, ce qui augmente le seuil ΔK_{th}. Tandis que la fragilisation par l'hydrogène et les autres mécanismes de corrosion classiques tendront à abaisser le seuil ΔK_{th} (Fatemi et Socie, 1988; Miller, 1987; Taylor, 1989). Les sous-chargements[1] « *underloading* » peuvent aussi causer temporairement une accélération de la propagation des fissures, contrairement aux sur-chargements *« overloading »* qui peuvent décélérer leur propagation (Dabayeh et Topper, 1995; Pompetzki, Topper et DuQuesnay, 1990; Taylor, 1989).

Les effets secondaires du soudage ne se limitent pas seulement au ΔK_{th} mais affectent aussi la limite d'endurance $\Delta\sigma_0[\mathrm{MPa}]$. En effet l'endurance qui est définie pour une durée de vie de N cycles, est influencée par : le fini de surface, l'épaisseur de la pièce, la fiabilité définie lors de la conception et les concentrations de contraintes; qui sont tous des paramètres affectés par les opérations de réparation et de soudage. Des modèles de correction ont été proposés par différents chercheurs pour tenir compte de l'effet du soudage dans les équations de

[1] Ces changements de régime accompagnés de changement dans les conditions thermo-hydrauliques favorisent l'apparition de la cavitation et de la corrosion.

propagation des fissures (Baik, Yamada et Ishikawa, 2011; Suresh, 1983) et pour la correction de la limite d'endurance (Therriault et Bernard, 2013). Cependant, à notre connaissance, ces corrections restent 'partielles' puisqu'elles ne tiennent pas compte de la nature aléatoire des différents facteurs et aussi de l'effet des différences entre le matériau de soudage et le matériau de base.

De cette revue, il ressort que la propagation de la fissure ne dépend pas seulement de sa localisation dans la turbine, sa géométrie et sa taille (Castillo *et al.*, 2010; Kitagawa, Yuuki et Ohira, 1975; Sanford, 2003; Xiong et Shenoi, 2008), mais aussi des conditions d'exploitation et des opérations de réparation. Par conséquent les données originales de la concepion des turbines ne peuvent être utilisées tout le long de la vie utile de l'équipement et d'où le besoin de les mettre à jour particulièrment le seuil de propagation ΔK_{th} et de l'endurance $\Delta\sigma_0$, pour une estimation plus précise de la fiabilité. Finalement, les réparations des cavitations et des corrosions engendrent aussi des changements dans les propriétés des matériaux et tenir compte de leurs historiques respectifs s'avère une étape incontournable dans le calcul de la fiabilité en fatigue.

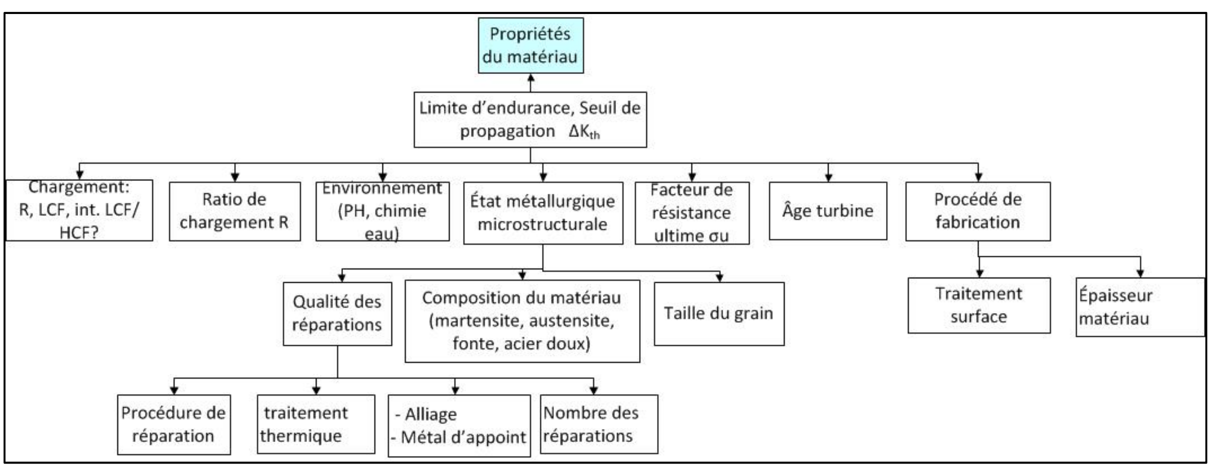


Figure 1.2 Facteurs affectant les propriétés du matériau

1.3 Inspections des fissures

L'inspection des turbines hydrauliques s'effectue généralement à l'usine avant leur mise en service. Des inspections ultérieures, typiquement après 50 ou 60 ans d'exploitation, peuvent avoir lieu pour déterminer l'état de fissuration de leurs aubes; dans ces cas, on parle d'inspection *in situ*. La détection et le suivi de ces fissures sont effectués moyennant des inspections périodiques qui peuvent être visuelles ou basées sur des techniques non destructives (NDT) telles que la méthode de magnétoscopie, « *Magnetic Particule Inspection (MPI)* », l'ultrason et le courant de Foucault (Onoufriou, 1999; Wall, Burch et Lilley, 2009). Ces techniques fournissent des informations importantes pour les calculs de fiabilité (Lassen, 2013; Veritas, 1992; Verma, Ajit et Karanki, 2010); inopportunément dans le cas des turbines hydrauliques, le plus souvent ces inspections sont visuelles, donc subjectives et peu précises et par conséquent, la confirmation de la probabilité de détection des fissures demeure extrêmement difficile à confirmer. Dans ces cas, on parle d'une probabilité de détection (POD *« Probability Of Detection »*) pour une taille donnée de fissure sur la surface de l'aube. La courbe de POD est définie en fonction de la taille de la fissure (a[mm]) et de l'incertitude reliée à l'outil de détection (Grooteman, 2008). Plusieurs formes de courbes de détection sont proposées dans la littérature, en fonction du cas et des méthodes d'inspection utilisées (Moan, 2005; Onoufriou, 1999; Rizzo, 2007).

Dans le cadre des turbines hydrauliques d'Hydro-Québec, l'inspection n'est pas effectuée d'une manière continue, car c'est une opération qui nécessite l'arrêt du groupe. Aussi, et à cause de l'incertitude inhérente à la méthode d'inspection (probabilité de détection), la taille d'une fissure à un instant *t* doit impérativement être considérée comme une variable aléatoire, ce qui exclue toute approche déterministe dans l'estimation de la fiabilité résiduelle en fatigue.

1.4 Élicitation des avis des experts

Dans certaines situations, le processus d'observation s'avère imparfait ou très couteux, c'est le cas dans l'évaluation de la taille du défaut ou dans l'évaluation des contraintes de chargement sur les aubes des turbines hydrauliques. Dans ces conditions et pour estimer les contraintes dues au chargement, Johannesson a proposé une approche permettant l'extrapolation temporelle sur une grande période, d'un historique réel de chargement enregistré sur une courte période (Johannesson, 2006). Ces travaux ont constitué un point de départ pour Szczota qui a proposé des générateurs de séquences synthétiques du chargement, utilisant les chaînes semi-Markoviennes et les méthodes de « *Bootstrap* » (Szczota, 2012). Récemment Diagne a proposé une méthodologie pour prédire le niveau des contraintes sur les aubes, à partir de mesures indirectes collectées *in situ,* (Diagne, 2016).

Pour l'estimation des tailles des fissures dans les turbines hydrauliques et à défaut de modèles physiques (ou empiriques), le recours aux avis des experts devient une alternative incontournable (Aven et Guikema, 2011; Clemen et Winkler, 1999). Dans ces cas, l'information peut être obtenue selon un processus d'élicitation visant l'obtention des données requises à travers des questions spécifiques (Ayyub, 2001). À ce niveau, il convient de distinguer entre le processus de l'élicitation et les techniques d'élicitation. En effet le processus d'élicitation décrit le déroulement de l'élicitation, le choix des experts, le rôle des facilitateurs, etc. Les techniques d'élicitation encadrent l'expert et décrivent, entre autres, les étapes de l'élcitation, le format de l'information, etc.

La Figure 1.3 représente un résumé succinct de la revue de littérature relative aux avis des experts.

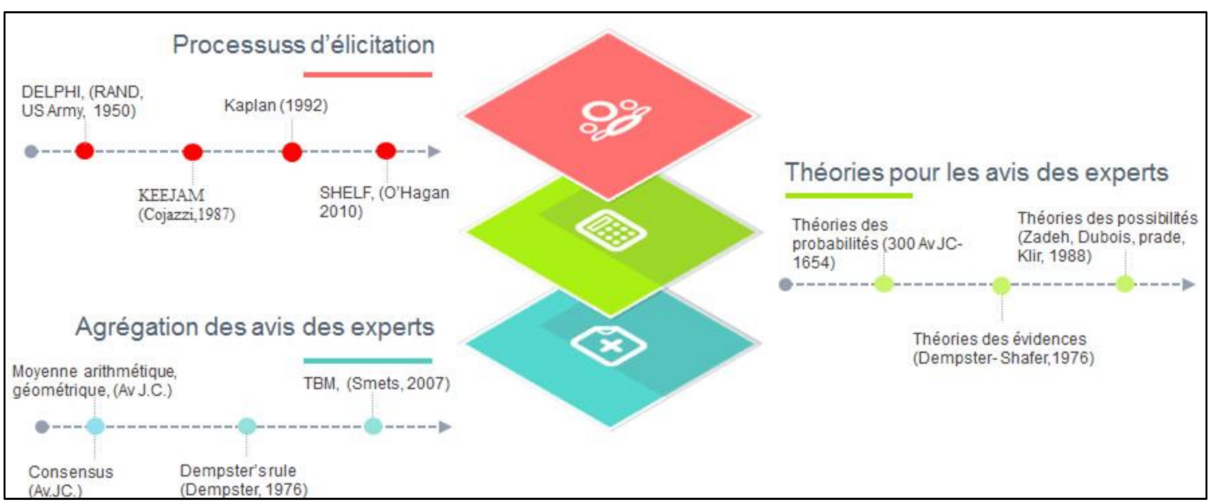


Figure 1.3 Sommaire succinct de la revue de littérature relatives aux avis des experts

1.4.1 Processus d'élicitation

Les protocoles d'élicitation proposés dans la littérature partagent généralement les mêmes principes. Certains se distinguent des autres par rapport au processus de déroulement d'élicitation, ou dans la technique de fusion et le rôle des facilitateurs et intégrateurs. Par exemple dans l'approche SHELF le choix du type de distribution découlant de la fusion des avis des experts est effectué par le facilitateur qui devrait aussi valider son choix en se procurant des avis supplémentaires des experts et en effectuant des analyses de sensibilité (O'Hagan, 2012).

L'approche la plus référée dans la littérature est l'approche Delphi; elle fournit un guide pour le choix des facilitateurs, le choix des experts ainsi que pour l'élaboration des questions. Les avis collectés selon cette approche, sont présentés sous forme de valeurs médianes et d'intervalle interquartile (soit les percentiles 25 % et 75 %), puis ils sont examinés par les experts et révisés, ultérieurement jusqu'à l'atteinte d'un consensus. À la fin du processus, un résumé est préparé pour documenter les différentes étapes du processus (Ayyub, 2001). La méthode Delphi reste une bonne approche pour atteindre un consens, par contre, à notre avis, il y a un manque sur les directives de son utilisation, ce qui peut nuire à la fiabilité des résultats atteints.

L'approche Kaplan, contrairement à Delphi, met l'emphase sur les connaissances des experts et non sur les experts eux-mêmes. En effet dans cette approche on demande aux experts d'exprimer leurs connaissances et expériences en lien avec le sujet en cours de traitement (Kaplan, 1992). Cependant, on reproche aux deux méthodes de ne pas favoriser l'interaction entre les experts; un aspect que l'approche de NUREG a tenté de combler. Pour cet effet l'approche NUREG (approche orientée principalement au domaine nucléaire) met l'accent sur la formation et les rencontres entre experts, qui sont choisis avec des profils diversifiés (NUREG-1150, 1990).

Dans la littérature on retrouve aussi l'approche KEEJAM où l'agrégation des avis des experts est basée sur l'approche bayésienne (Cojazzi et al., 1987). On retrouve aussi l'approche SEPT (*Standard for Educational and Psychological Testing*) qui fournit des guides pour la construction des tests, leurs évaluations, la formulation des critères d'évaluation des tests ainsi que leurs effets (Ayyub, 2001). On peut citer aussi l'approche de Truong *et al* (Truong et Heuvelink, 2013) et l'approche des scénarios (Ayyub, 2001; Ha-Duong, 2008).

Chacun de ces processus d'élicitation a été bâti de façon à rencontrer des exigences précises, dictées par le contexte d'élicitation , le mode de gestion du processus d'élicitation, la nature des questions, la nature de l'information attendue (consensus ou autre) et le rôle du facilitateur. Ces contextes ne représentent pas forcément les mêmes orientations visées par notre projet, d'où le besoin d'élaborer un processus d'élicitation, adapté aux objectifs du projet.

1.4.2 Techniques d'élicitation

Les techniques d'élicitation sont des encadrements visant à guider et supporter les experts à exprimer leurs avis, qui dans certains cas doivent avoir un format assez particulier pour répondre aux exigences du système étudié. Dans la littérature, les techniques d'élicitation sont généralement classées en deux catégories: directes ou indirectes. Dans la catégorie des techniques directes, les experts expriment leurs jugements en répondant directement aux questions posées. Dans ce sens, on cite les techniques basées sur: la probabilité, la distribution de probabilité et la théorie bayésienne (Meyer et Booker, 2001; O'Hagan, 2012; Rocquigny, 2012). Contrairement aux techniques directes, dans les techniques indirectes on suppose que l'information recherchée n'est pas facilement accessible. Dans ces cas on utilise des approches basées sur l'évaluation des similitudes et des différences entre certains concepts pour aboutir à l'information requise (Hudlicka, 1996). Cette mesure de similarité (proximité) entre concepts, constitue une mesure relative qui est généralement liée à un contexte spécifique (Cooke, 1994). Par conséquent, l'adoption de ces techniques devrait être utilisée avec précaution car l'expert peut avoir du mal à comprendre le but des questions en raison du manque de transparence dans le processus; ce qui pourrait affecter la qualité de la réponse (Hora, 1992). De ce fait, cette classe de techniques devrait être adoptée dans les cas où les quantifications directes sont impossibles (Cooke et Goossens, 1999), surtout que cette catégorie de techniques nécessite plus de temps pour l'analyse des concepts (Cooke, 1994).

D'autre part, les experts ne font pas nécessairement tous partie de la même *école de pensées*, ce qui se traduit souvent par une différence significative dans leurs avis. Pour cerner ces dispersions, certains chercheurs recommandent l'adoption des approches telles que l'approche du « *vocabulaire calibré* » conjointement avec une échelle de probabilité (Fallet *et al*., 2011).

De cette revue, il découle que le choix d'une technique appropriée pour l'élicitation des avis des experts dépend impérativement du domaine d'application et de la forme requise de l'information. Par conséquent, il sera intéressant de comparer l'effet de l'utilisation des différentes techniques, pour en ressortir, si possible, celles qui sont plus adaptées pour la prédiction des valeurs dans un contexte sans historique ou valeurs de référence.

1.4.3 Théories de modélisation des avis des experts

Dans certains cas, les données élicitées doivent être exprimées sous une forme spécifique pour répondre aux exigences des systèmes qu'elles représentent. Ainsi pour les cas nécessitant des estimations simples, Meyer et Brooker suggèrent leur représentation avec une valeur de probabilité et pour les systèmes nécessitant plusieurs estimations (un ensemble d'estimations), l'avis de l'expert pourrait être exprimé avec des distributions de probabilité (Meyer et Booker, 2001). Le choix de la théorie des probabilités a été aussi supporté par O'Hagan qui préconise aussi l'utilisation de la théorie des probabilités classique puisque les analystes et les experts sont généralement plus familiers avec cette théorie (O'Hagan, 2012). Cependant son inconvénient réside dans sa limitation dans la représentation des données et dans les opérations arithmétiques. Par exemple, l'opération $P(a \cup b) = P(a) + P(b)$ n'est exacte que si a et b sont deux événements disjoints. En effet dans cette expression on ne tient pas compte de l'interaction mutuelle des deux évènements (a, b) ce qui peut mener, dans certaines conditions, à un résultat différent de $P(a) + P(b)$. Afin d'éviter ces limitations, certains chercheurs ont proposé l'utilisation de la théorie de la logique floue combinée avec certaines propriétés probabilistes pour modéliser l'incertitude des données épistémique (Möller et al., 2001). D'autres chercheurs ont recommandé des encadrements basés sur la théorie des intervalles (Limbourg et De Rocquigny, 2010), tandis que certains ont opté pour la théorie des évidences car elle permet de prendre en compte les effets mutuels entre les événements et permet ainsi la combinaison de données provenant d'autres théories, telles que la probabilité pure et la théorie des possibilités (Destercke, Dubois et Chojnacki, 2006; Klir et Smith, 2001). Cette propriété, de la théorie des évidences, est due à sa capacité à modéliser l'incertitude épistémique (l'ignorance) et de la refléter directement dans des fonctions d'appartenance (Baraldi, Popescu et Zio, 2010). De plus, ces caractéristiques sont maintenues le long du processus de combinaison, ce qui n'est pas le cas avec l'approche probabiliste classique où l'ignorance n'est pas considérée (Yager et Liu, 2008).

L'effet de la diversification des théories de modélisation et de leurs différences dans la représentation des informations, sur le calcul de la fiabilité est à notre connaissance, méconnu et n'a été étudié que très partiellement. À cet égard, on cite le travail de Baudrit *et al*. qui ont comparé trois théories: *P-B*ox, la théorie des intervalles et la théorie des possibilités. Ils ont conclu que la théorie des possibilités était la plus précise dans la description de l'information (Baudrit et Dubois, 2005). Beer *et al*. avancent que le choix d'une représentation particulière de l'information dépend surtout du niveau de risque toléré (Beer, Ferson et Kreinovich, 2013). Flage *et al*., ont comparé trois types de représentation selon : la théorie des probabilités, la théorie des possibilités et selon une théorie hybride qui combine la théorie des probabilités avec la théorie des possibilités. Leur étude a montré que les trois approches sont comparables et que les résultats de l'approche hybride sont délimités par les résultats de la théorie des probabilités et de la théorie des possibilités (Flage et al., 2013).

Cette diversification dans les conclusions des différentes recherches soulève la question sur l'effet du domaine d'étude et des techniques d'élicitation sur les avis des experts. Ce qui suggère une étude plus approfondie des techniques d'élicitation afin de ressortir une technique qui soit plus compatible à la nature de notre projet de recherche.

1.4.4 Agrégation des avis des experts

La motivation derrière l'utilisation de plusieurs experts découle du désir d'obtenir le maximum d'informations possibles (Clemen et Winkler, 1999). Dans ce cas, l'agrégation des avis des experts peut s'effectuer selon différentes approches : des approches comportementales ou des approches mathématiques. L'agrégation par consensus en est un exemple d'agrégation comportementale. Dans ce cas le facilitateur cherche à atteindre un accord (global ou partiel selon les cas) entre les différents experts. L'avantage de cette approche c'est qu'elle permet de renforcer les liens sociaux, mais en contrepartie, les résultats qui en découlent peuvent contenir des biais dû au profil des membres ou à la dominance de certains sous-groupes. De ce fait, les agrégations selon des approches mathématiques seront plus préférées. Dans ce sens, on cite les agrégations basées sur: les moyennes pondérées, la règle de Dempster, ou bien la mesure d'entropie de Shannon (Ayyub, 2001). À titre d'exemple la règle de Dempster est particulièrement intéressante puisqu'elle permet de combiner des données pouvant provenir de différentes théories. Les méthodes basées sur la moyenne (pondérée ou pas) ont une autre particularité. En effet, elles donnent plus de poids aux opinions exprimées par le plus grand nombre d'experts, ce qui risque de masquer la divergence entre experts; une divergence qui peut prendre dans certains cas la forme d'une contradiction entre les avis des experts et par suite les méthodes d'agrégation basées sur l'intersection des opinions ne peuvent pas s'appliquer puisque cette intersection est vide (Ha-Duong, 2008). Il y a environ 10 ans, Smets a présenté la théorie de TBM « *Transfer Belief Model* » comme une solution à cette problématique et comme recommandation, Smets suggère de résoudre les conflits à leurs racines en imposant un système approprié pour réduire les conflits (Smets, 2007).

À la lumière de nos lectures nous pouvons avancer que globalement la problématique de fusion des opinions conflictuelles n'est pas totalement cernée et par conséquent, beaucoup d'efforts devront être déployés pour atteindre des résultats acceptables. D'un autre côté, les approches de combinaison sont, par essence, sensibles au nombre des experts (Kaplan, 1992), à leurs formations, à leurs expériences, à l'interaction avec les autres experts et aux règles de combinaison, qui devront être flexibles et appropriées au problème étudié. Dans ce cas, une analyse de sensibilité nous semble conseillée pour valider le choix de l'approche et les règles appropriées.

Finalement, la crédibilité et la confiance dans les avis des experts est un facteur aussi important que le choix des règles d'agrégation puisqu'il affecte la qualité des jugements. Il donc est impératif d'en tenir compte lors des fusions des avis des experts (Aven, 2012; Aven et Guikema, 2011).

1.5 Conclusion

Lors de cette revue nous avons exploré les phénomènes intervenant dans le processus de fissuration par fatigue et qui affecte aussi les propriétés mécaniques du matériau des turbines et par conséquent le calcul de fiabilité. Ainsi pour une estimation plus précise de la fiabilité, il est primordial de mettre à jour ces paramètres en faisant appel aux avis des experts. L'éventail de littérature disponible sur les processus d'élicitation des avis des experts a permis de constater que le besoin en avis des experts n'est pas exclusive au domaine de fiabilité en fatigue, mais qu'il est commun à d'autres domaines tels que les secteurs: médical, de technologie de l'information, marketing, etc. Cependant, les techniques d'élicitation proposées dans la littérature ont été élaborées pour répondre à ces besoins spécifiques. Pour le domaine de la fiabilité en fatigue, notre revue de littérature a révélé, qu'aucune comparaison des techniques d'élicitation impliquant des modèles complexes comme celui de Gagnon *et al.*, n'a été entamée auparavant. D'où la nécessité de définir des techniques d'élicitation appropriées, au contexte de notre projet de recherche.

Les points soulevés dans cette revue, associés à l'usage des probabilités imprécises et des techniques d'élicitation dans la problématique de fiabilité en fatigue, n'ont pas fait l'objet (à notre connaissance) de publications auparavant et seront traités dans le cadre de cette thèse pour la première fois, dans l'objectif de ressortir les pratiques qui s'adaptent le mieux avec le projet de fiabilité en fatigue.

CHAPITRE 2

IMPRECISE PROBABILITIES IN FATIGUE RELIABILITY ASSESSMENT OF HYDRAULIC TURBINES

Mounia Berdai[1], Antoine Tahan[1], Martin Gagnon[2]

[1]École de technologie supérieure, 1100, rue Notre-Dame Ouest, Montréal (Québec) H3C 1K3 Canada
[2]IREQ, 1800 Boulevard Lionel-Boulet, Varennes (Québec) J3X 1S1, Canada

This chapter has been published in the "*ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems"* Part B: Mechanical Engineering, Published online: November 21, 2016, Vol. 3, Issue 1 (March 2017).

2.1 Abstract

Risk analyses are often performed for economic reasons and for safety purposes. In some cases, these studies are biased by epistemic uncertainties due to the lack of information and lack of knowledge, which justifies the need to expert opinion. In such cases, experts can follow different approaches for the elicitation of epistemic data, using probabilistic or imprecise theories. But how do these theories affect the reliability calculation? What are the influences of using a mixture of theories in a multi-variable system with a non-explicit limit model? To answer these questions, we propose an approach for the comparison of these theories, which was performed based on a reliability model using the FORM (First order reliability method) approach and having the Kitagawa-Takahashi diagram as limit state. We also propose an approach, appropriate to this model, to extend the reliability calculation to variables derived from imprecise probabilities. For the chosen reliability model, obtained results show that there is a certain homogeneity between the considered theories. The study concludes also that priority should be given to expert opinions formulated according to unbounded distributions, in order to achieve better reliability calculation accuracy.

Keywords: Uncertainty, reliability, epistemic, expert opinion elicitation, bounded distribution, nonlinear limit state, Hasofer-Lind, Kitagawa-Takahashi

2.2 Introduction

The reliability of hydroelectric turbines is a complex function which depends mainly on: turbine age, cumulative damage, design data, loading cycles, periodic inspections and material intrinsic properties. Based on a probabilistic approach, Gagnon *et al.* have proposed a reliability model of hydroelectric turbine blades (Gagnon et al., 2013), having as limit state function $g(x)$, which is determined by the Kitagawa-Takahashi diagram and has as main inputs: the loading stress ($\Delta\sigma$ – [MPa]) and the defect size ($a -$ [mm]).

The limit state function $g(x)$ is mainly influenced by the material mechanical properties, which depend on fabrication process and residual stress resulting from the welding process (Thibault, Gagnon et Godin, 2015).Thus the original design values cannot be used during the entire useful life of the hydraulic turbine and therefore it is important to update such characteristics.

Also Up-to-date assessments of defect size and loading stress are often difficult to obtain, due to hydraulic turbine operational constraints; which suggest the use of expert opinion to fill this lack of information. In such situation, experts can express their opinions by providing prior probabilities for the required data, or update it by using a Bayesian approach (Aven et al., 2013; Ayyub, 2001). The question of how to incorporate expert opinion in the reliability index calculation of such model was not addressed before and deserves to be assessed.

To express these expert opinions, O'Hagan advocated the use of classical probability theory because generally analysts and experts are generally more familiar with such theories (O’Hagan, 2012). Some researchers proposed the use of Fuzzy logic theory combined with probabilistic properties to model the epistemic data uncertainty (Möller et al., 2001), while others recommend using the interval theory (Limbourg et De Rocquigny, 2010). In this sense, Ayyub distinguishes two types of intervals: 'crisp interval' and 'fuzzy interval' (Ayyub, 2001). Arithmetic operations on to these intervals, handle the lower and upper interval bounds; which don't necessarily inform us on the interval behaviour between the two bounds. As an alternative, the possibility theory seems to be a good choice. In this theory, the experts can express their opinion under a belief function by assigning a belief degree to each element of the domain. The possibility theory was proposed by Zadeh in 1978 as an extension to the fuzzy set theory (Zadeh, 1965; Zadeh, 1978) and its use has been spread mainly through the work of Dubois, Prade and Smets (Dubois et al., 2004; Dubois et Prade, 1998; Dubois, Prade et Smets, 2008). Furthermore, evidence theory is another theory based on belief function, which can be also used for expert opinion modelling. This theory has the advantage of taking into account the mutual effects between events and allows the combination of data from other theories, such as pure probability and the theory of possibilities (Destercke, Dubois et Chojnacki, 2006; Klir et Smith, 2001). This characteristic is due to its ability to model epistemic uncertainty and ignorance and to reflect them directly in membership functions (Baraldi, Popescu et Zio, 2010). More, these data characteristics are preserved along the combination process, which is not the case with the probabilistic approach, where ignorance is not considered (Yager et Liu, 2008).

The diversification of the discussed theories and the difference in their respective information representations are the source of our motivation to compare them in order to assess their impact on the reliability calculation. In this regard, Baudrit *et al.* compared three techniques: $P-box$, random set, and possibility theory and they concluded that the possibility theory was the most precise one (Baudrit et Dubois, 2005). Beer *et al.* compared some approaches from the imprecise probability family and they concluded that the choice of the information representation model depends mostly on the risk level. They showed that the interval approach is particularly interesting due to its conservatism which can highlight important consequences of extreme (less frequent) events (Beer, Ferson et Kreinovich, 2013). This conclusion is mostly suitable for specific domains, but cannot be generalized to cases where we are looking for an accurate risk calculation. Flage *et al.*, conducted another type of comparison between representations based on, pure probability theory, possibility theory and hybrid representation (Flage et al., 2013). Their analysis showed that the hybrid approach results are bounded by the results from the two other approaches. Also, they showed that hybrid approach has the advantage of maintaining the 'uncertainty' formulation during the propagation process (Flage et al., 2013).

Results from previous studies provide interesting finding, regarding the use of some imprecise probability but do not give complete answers to our main motivation expressed at the beginning of this section; hence the idea of extending this comparison to theories such as: intervals, possibility and evidence theories. Our goal is to evaluate the behavior of these theories on a propagation model more complex than the one presented by other authors (Flage et al., 2013).

The effectiveness of the imprecise theories (possibility, evidence) depends mainly on the choice of the membership functions (or belief functions), the selection of experts and their awareness on these theories (Beer, Ferson et Kreinovich, 2013; Masson, 2005). In fact, neglecting any factor can lead in some cases to results disparities (Ferson et al., 2004); an aspect which we intend also to explore in this study.

This paper is structured as follows: Section 2.3 provides the methodology followed in this paper. Section 2.3.4 and 2.3.5 treat some specific processing for the studied model. Finally, Sec. 2.3.6 discusses the achieved results.

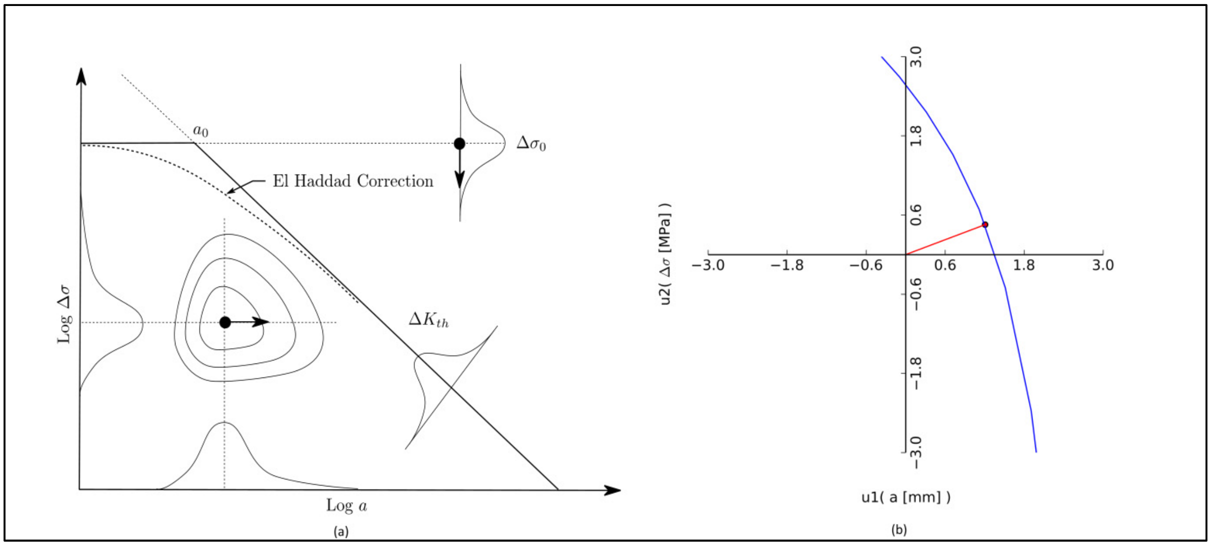


Figure 2.1 (a) Reliability model based on Kitagawa–Takahashi diagram and (b) reliability index in the iso-probabilistic space

2.3 Methodology

The purpose of this paper is to assess the impact of different modeling theories, on a probabilistic reliability model and to evaluate their effects on a system governed by a nonlinear limit state. We then assess the impact of a theory mixture on reliability calculations. This situation may correspond to a case where the information is derived from several experts and is expressed under various theories, hence the proposal of a case matrix with different sensitivity cases. The comparison of different theories is performed through the comparison of the resulting reliability index, knowing that these theories have equivalent support domains. The reference case corresponds to the case where data is represented by a Gumbel distribution.

The proposed methodology consists in the following steps:

- Generating case matrix in order to perform comparison between different modelling theories. During this step we assume that the endurance limit σ_0 and propagation threshold ΔK_{th} remain constant and similar for all proposed cases, thereby ensuring the same limit state function $g(x)$ for all proposed cases.

- Defining the support domain and the associated parameters for each suggested theory in the case matrix. The purpose of this step is to ensure that all theory support domains are equivalent, in order to reduce disparity between them. Section 2.3.2 suggests some approaches on how to define these supports domains.
- Adapting 'bounded theories', proposed in the case matrix, to Hasfoer-Lind approach requirements.
- Processing of suggested imprecise theories and converting imprecise probability to probability distribution.
- And performing reliability index calculations, based on the model proposed by Gagnon *et al.* (Gagnon et al., 2013), which uses Hasofer-Lind approach and Kitagawa-Takahashi diagram as a limit state.

2.3.1 Case study

To perform the comparison between different modeling theories, we use the reliability model based on Hasofer-Lind approach and having as a limit state function $g(x)$, determined by the Kitagawa-Takahashi diagram (Gagnon et al., 2013). This model is a multi-variable and non-linear model with a non-linear limit state function $g(x)$. This limit state is composed of two thresholds: the first threshold is the stress intensity range for crack growth $\Delta K_{th}[\mathrm{MPa}\sqrt{\mathrm{m}}]$ where $\Delta K_{th}$is defined by Linear Elastic Fracture Mechanics (LEFM), and the second threshold is the fatigue limit $\Delta\sigma_0$ [MPa] (Figure 2.1(a)).The space under these two limits is known as the 'safe' operation region. In this region we assume that there is no-crack propagation (or a very slow propagation). In fact, a flaw will become a detectable crack after the HCF (High cycle fatigue) onset which marks the point in time for crack growth (Gagnon et al., 2013).

The main inputs parameters of this model are: the loading stress ($\Delta\sigma$ – [MPa]) and the defect size ($a -$ [mm]) (Figure 2.2), which determine the reliability index, based on the Hasofer-Lind index approach, as shown in Figure 2.1 (b). This approach allows also the calculation of the reliability index β_{HL} and the failure probability P_f, based on a linear approximation of the limit state function (First Order Reliability Methods: FORM) and the input variable probability distribution. The failure probability is given by the following equation: $P_f \approx 1 - \varphi(\beta_{HL}) = \varphi(-\beta_{HL})$ where φ is the cumulative function of a standard Gaussian distribution. This reliability probability does not mean that a turbine blade is in a failure mode; but it indicates the position of the operation point regarding the safe region.

2.3.2 Case matrix

In the reliability model proposed by Gagnon *et al.* (Gagnon et al., 2013), loading stress and defects size are the main input variables. These two random variables represent the extreme values in a given volume element and should be described by an extreme value theory (EVT) using Weibull or Gumbel distributions. The choice of the appropriate distribution characterizing a ' random variable' should be consistent with the physical system behavior and with the range values of interest (Castillo et al., 2014), hence the choice of Gumbel distribution which has already been used previously to model extreme loading stress and defect size (Gagnon et al., 2013). Later in Section 2.3.3, this distribution will be considered as a reference distribution.

On the other hand, the assessment of such input parameters is often difficult due to hydraulic turbine operational constraints, which justify the need to expert opinion to fill this lack of information. The simplest way to express this opinion would be to formulate it according to interval theory. In this theory only the upper and lower interval bounds are manipulated during the propagation process (Ayyub, 2001), which could be inappropriate for nonlinear models with a non-monotonic transfer function, because in this case we could miss the extreme points. Therefore, it’s important to consider all elements in the interval, during the propagation step. To fulfill this objective, most elements in the interval should be taken into account, which might result in assuming that all the interval elements follow a uniform probability distribution. Hence the choice of a uniform probability distribution as sensitivity case in the case matrix (Tableau 2.1). This hypothesis has also been used by Beer *et al.* (Beer, Ferson et Kreinovich, 2013).

In other situations, the expert might have more details to describe the information which allows him to express it under a belief function by assigning a belief degree to each element of the support domain. This could result in formalized possibility distributions or evidence distributions, depending on the followed approach. Possibility distribution can have several shapes; which will affect differently the propagation result, depending on the input variable distribution. References (Zadeh, 1965; Zadeh, 1978) give more details on possibility theory.

Generally, possibility distribution is often modeled with a triangular shape (Aven et al., 2014; Dubois et al., 2004), so it would be interesting to compare this distribution with a triangular probability distribution, as suggested in the case matrix (Tableau 2.1).

With these cases, we attempt to compare formulations based on classical probability distributions (e.g. uniform, Gumbel and triangular distribution) against the formulations based on imprecise probabilities (possibility and evidence) (Figure 2.3).

Thus in the proposed case matrix, *loading stress range* (Δσ – [MPa]) and *defect size* ($a -$ [mm]), can follow five behaviors: Gumbel distribution (F_G), uniform distribution (F_U), triangular distribution (F_T), evidence distribution and a possibility distribution (Tableau 2.1). We assume that in all proposed cases, the material properties remain constants, which lead to a single limit state function; thereby the case matrix will be reduced to 25 cases for each point chosen in the probabilistic space, as shown in Figure 2.4.

Tableau 2.1 Matrix case with the 25 combinations corresponding to different defect size (a) and loading stress ($\Delta\sigma$) distributions. F_U is the uniform distribution, F_T is the triangular distribution, F_G is the Gumbel distribution, P is the possibility theory and E is the evidence theory

$a \sim F_U, \Delta\sigma \sim F_U$	$a \sim F_G, \Delta\sigma \sim F_U$	$a \sim F_T, \Delta\sigma \sim F_U$	$a \sim E, \Delta\sigma \sim F_U$	$a \sim P, \Delta\sigma \sim F_U$
$a \sim F_U, \Delta\sigma \sim F_G$	$a \sim F_G, \Delta\sigma \sim F_G$	$a \sim F_T, \Delta\sigma \sim F_G$	$a \sim E, \Delta\sigma \sim F_G$	$a \sim P, \Delta\sigma \sim F_G$

$a \sim F_U, \Delta\sigma \sim F_T$	$a \sim F_G, \Delta\sigma \sim F_T$	$a \sim F_T, \Delta\sigma \sim F_T$	$a \sim E, \Delta\sigma \sim F_T$	$a \sim P, \Delta\sigma \sim F_T$
$a \sim F_U, \Delta\sigma \sim E$	$a \sim F_G, \Delta\sigma \sim E$	$a \sim F_T, \Delta\sigma \sim E$	$a \sim E, \Delta\sigma \sim E$	$a \sim P, \Delta\sigma \sim E$
$a \sim F_U, \Delta\sigma \sim P$	$a \sim F_G, \Delta\sigma \sim P$	$a \sim F_T, \Delta\sigma \sim P$	$a \sim E, \Delta\sigma \sim P$	$a \sim P, \Delta\sigma \sim P$

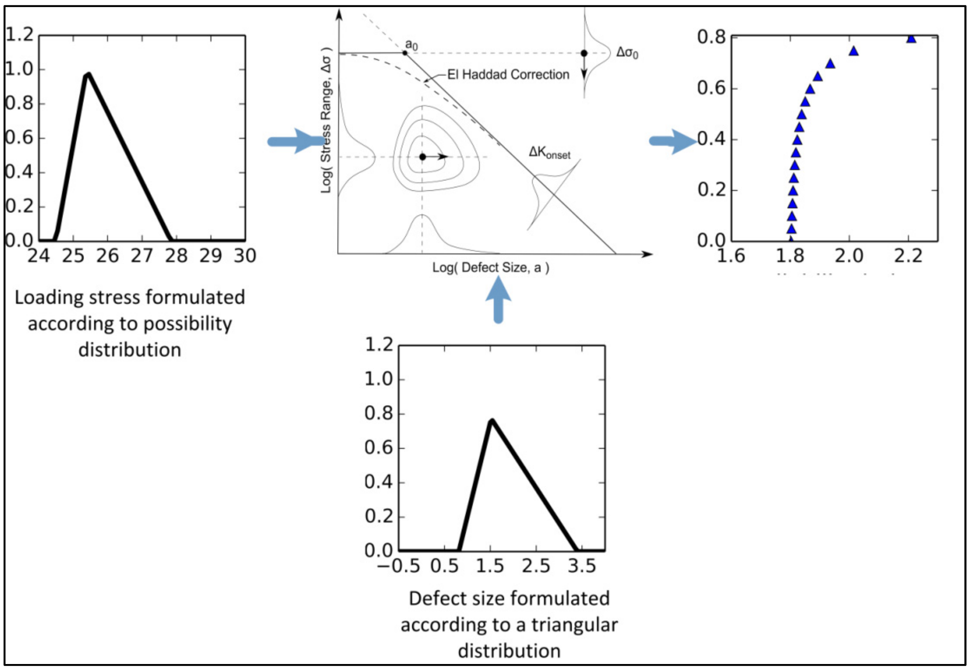


Figure 2.2 Example of the reliability index distribution, resulting from loading stress following a possibility distribution and defect size formulated according to a triangular distribution

2.3.3 Distribution support

As mentioned, the comparison between the proposed theories is performed through the reliability index calculation. For this purpose, the suggested theories should have equivalent support domains (Tableau 2.1), in order to reduce results disparity. Therefore, from a reference distribution provided by an expert, we can determine the lower and upper bounds of the support domain delimiting an area of interest. And knowing these support bounds we can also deduce the characteristics of another distribution, based on its mathematical cumulative density function (CDF) expression. The determination of the lower and upper support bounds could be performed, according to different approaches such as: the quantile method and the approach based on the Bienyamé-Tchebechev inequality (Dubois et al., 2004).

The proposed quantile approach consists in:

- Step 1: Calculating quantiles $x_{\frac{\alpha}{2}}$ and $x_{1-\frac{\alpha}{2}}$ corresponding respectively to $\frac{\alpha}{2}$ and $(1-\frac{\alpha}{2})$ of the cumulative density function (CDF) for the reference distribution F_{D_r}.

$$F_{D_r}\left(x_{\frac{\alpha}{2}}\right) = \frac{\alpha}{2} \text{ and } F_{D_r}\left(x_{1-\frac{\alpha}{2}}\right) = 1 - \frac{\alpha}{2} \tag{2.1}$$

- Step 2: Determining the characteristics of the cumulative density function F_D (theory to be compared to the reference distribution), based on its CDF mathematical expression, the reference distribution characteristics and quantiles found in the previous step ($x_{\frac{\alpha}{2}}$ and $x_{1-\frac{\alpha}{2}}$), by using the following equation:

$$F_D\left(x_{\frac{\alpha}{2}}\right) = \frac{\alpha}{2} \text{ and } F_D\left(x_{1-\frac{\alpha}{2}}\right) = 1 - \frac{\alpha}{2} \tag{2.2}$$

Also in this approach, it is assumed that the CDF mode value of the distribution D is identical to CDF mode value of the reference distribution D_r (Figure 2.5).

The quantile approach could be replaced by the Bienyamé-Tchebechev approach which seems easier regarding its calculation process, but it provides conservative reliability indices; a fact that may mask the searched difference between theories. Hence, the adoption of the quantile approach, that can also be adopted to ensure the transition between an imprecise probability theory and a pure probability distribution (Section 2.3.5).

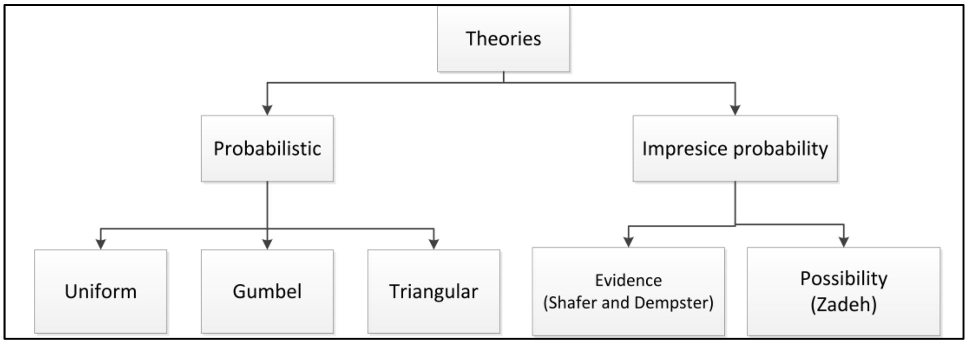


Figure 2.3 Theories used in the case matrix

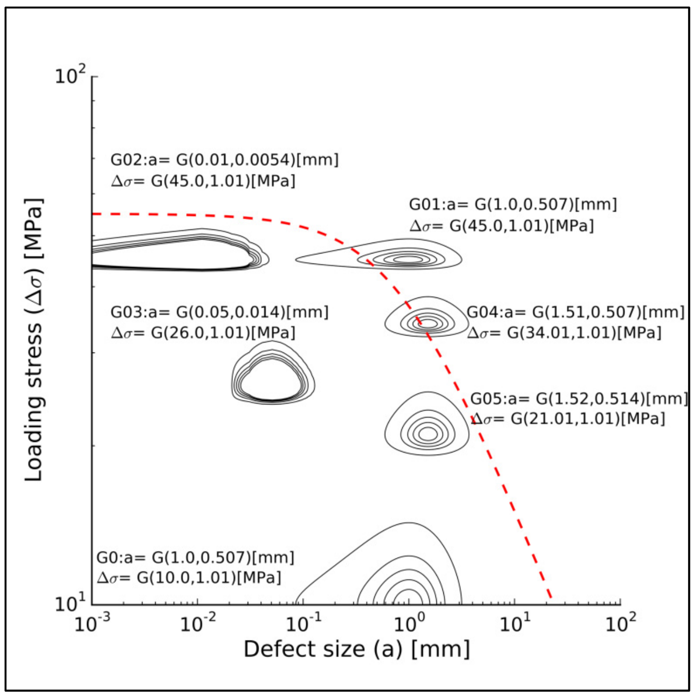


Figure 2.4 Points from the probabilistic domain used in the matrix case (Tableau 2.1). In this figure, defect size and loading stress are expressed with Gumbel distribution

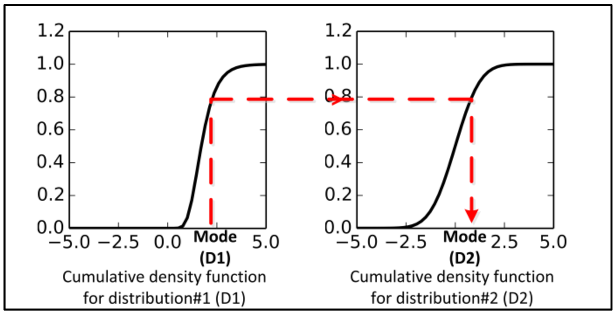


Figure 2.5 Example of determining the mode of distribution#2 (D2) from distribution#1 (D1)

2.3.4 Requirement of the reliability model

The Hasofer-Lind reliability index calculation, using Kitagawa-Takahashi diagram as a limit state, could be summarized to the calculation of the shortest distance z separating the iso-probability space origin and the system limit state $g(x)$. This distance can be written as: $z = \sqrt{u_1^2 + u_2^2}$ where u_1 is the standardized variable associated to the defect size (a) and u_2 is the standardized variable associated to the loading stress (Δσ) (Figure 2.1). The search of z optimum value (inside the iso-probabilistic domain) means that z has finite values and therefore u_1 and u_2. Thus a finite reliability index means that in the iso-probabilistic space we have at least one finite value for u_1 and u_2.

If the defect size (a) and the loading stress (Δσ) are described by bounded distributions and if the considered points of assessment are outside of their distribution support domain (i.e. their cumulative density functions are null or one), then the transformation into the iso-probabilistic space will be infinite, resulting in an infinite value of the reliability index β_{HL}. However, this reliability index would be finite if the defect size (a) and the loading stress (Δσ) were described by unbounded distributions. Therefore, the adoption of bounded distributions in the reliability calculation based on Hasofer-Lind approach seems influencing the results sensitivity. This impact becomes more important when the variable's uncertainty is small which reduce the support domain of the bounded distribution. This issue, of bounded distribution, has been addressed only by a few authors in the published literature. In Beck *et* *al.*(Beck, 2016), they suggest some alternatives for the design point search but these alternatives don't necessarily comply with the system physical behaviour (Castillo et al., 2014). To meet this requirement, bounded distribution should be converted to an unbounded one, suitable with the physical behaviour of the studied system. The result in this case will depend on the chosen conversion methodology and the 'tolerance' to the introduced subjectivity, which will alter the original information shape and thus affects the reliability calculation (see Section 2.3.5). As a solution, we propose to 'extend' the bounded distributions, such that having their tails identical to those of a chosen unbounded probability distribution P coherent with the system's physical behaviour (Castillo et al., 2014). By this way, we preserve the original 'data characteristic' as it was formulated by the expert.

So for a bounded probability distribution D characterised by a probability density function f_D with a lower and upper support bounds d_{inf} and d_{sup} respectively, the extension of this distribution according to a probability distribution P having as probability density function f_P, could be performed as follows:

$$f_{D-e}(x) = \begin{cases} f_P(x) & x < d_{inf} \\ f_D(x) & d_{inf} \leq x \leq d_{sup} \\ f_P(x) & x > d_{sup} \end{cases} \quad (2.3)$$

where $D - e$ represents the extended distribution.

2.3.5 Imprecise theories processing

The reliability calculation according to FORM approach requires that the input random variables follow a probability distribution. However, data can be expressed using other theories than probabilistic distribution. In such cases the possible options are: converting imprecise probability to probability distribution, converting imprecise probability by application of correction factors, or converting imprecise probability to probability distribution by using '$\alpha - cut$' sampling.

2.3.5.1 Converting imprecise probability

Converting data from imprecise probability to a probabilistic data requires applying conversions between theories, which should obey to some criteria such as: the specificity, consistency, information invariance, symmetry preservation and ignorance conservation. Oussalah has treated all these criteria in detail (Oussalah, 2000). However, the conversion from evidence theory to probability theory, did not receive the same interest and few studies have been performed in this sense, as the work of Cobb *et al.* who have proposed an approach based on the conversion of the plausibility distribution to probability distribution (Cobb et Shenoy, 2006).

The transformations discussed above (Cobb et Shenoy, 2006; Oussalah, 2000) do not necessarily lead to an empirical distributions or specific distribution shape, which comply with the propagation model and the system's physical behaviour (Castillo et al., 2014). Hence, the proposition of the transformation based on quantile approach as discussed in Section 2.3.3 which allows us to move from a possibility theory (for example) to a probability distribution from our choice, suitable with the system's physical behaviour. This approach can also be applied to move from a given probability distribution to another probability distribution of our choice. As an example, we convert a possibility distribution to a Gumbel distribution, according to the quantile approach (Section 2.3.3). In this case, it is assumed that the possibility distribution is expressed according to a triangular belief function, delimited by two bounds a_{min} and a_{max}, with a mode c. Its probability density function is given by the following expression:

$$f(x) = \begin{cases} \dfrac{x - a_{min}}{c - a} & a_{min} \leq x \leq c \\ \dfrac{a_{max} - x}{a_{max} - c} & c \leq x \leq a_{max} \\ 0 & otherwise \end{cases} \tag{2.4}$$

And its cumulative density function could be expressed as follows:

$$F_D(x) = \begin{cases} \dfrac{(x-a_{min})^2}{2(c-a_{min})} & a_{min} \le x \le c \\ \dfrac{1}{2(a_{max}-c)}[-x^2+2a_{max}x - a_{min}a_{max} + a_{min}c - a_{max}c] & c \le x \le a_{max} \\ \dfrac{(a_{max}-a_{min})}{2} & x > a_{max} \\ 0 & otherwise \end{cases} \tag{2.5}$$

Note that the final value of the possibility pseudo-cumulative density function is not necessarily 1 because the distribution is not a probability distribution.

Let $x_{\frac{\alpha}{2}}$ be the quantile value corresponding to a level of $\frac{\alpha}{2}$ and $x_{1-\frac{\alpha}{2}}$ the quantile value corresponding to a level of $(1-\frac{\alpha}{2})$. We wish to convert this distribution into a Gumbel distribution whose cumulative density function could be written as follows:

$$F_r(x) = e^{-e^{(\frac{\mu-x}{\beta})}} \tag{2.6}$$

Where μ represents the Gumbel localization factor and β its scale factor. The application of the approach described in Section 2.3.3 leads to $\mu = \frac{x_{\frac{\alpha}{2}}\left[Log(-Log\left(1-\frac{\alpha}{2}\right))\right]-x_{1-\frac{\alpha}{2}}\left[Log(-Log\left(\frac{\alpha}{2}\right))\right]}{\left[Log(-Log\left(1-\frac{\alpha}{2}\right))\right]-\left[Log(-Log\left(\frac{\alpha}{2}\right))\right]}$ and $\beta = \frac{\mu-x_{\frac{\alpha}{2}}}{\left[Log(-Log\left(\frac{\alpha}{2}\right))\right]}$. So Gumbel parameters were obtained from the possibility distribution.

Some researchers have opted for simplified conversions to approach imprecise distributions with probabilistic distributions by applying some correction factors (Lasserre, 1999); a practise which looks like a deviation of the conversion described previously.

Converting imprecise probability to probability distribution can be also achieved by using 'α-cut' sampling. This option aims to preserve the original imprecise distribution shape in the reliability model, by propagating the intervals resulting from its sampling according to '$\alpha - cut$' process; an approach which has been followed by many researchers (Baraldi, Popescu et Zio, 2010; Ferson et al., 2004). In the traditional approach, this propagation is performed by considering the lower and upper bounds of each sampled interval. Or if the propagation model is not monotone, then we risk missing the extrema of this interval image, hence the idea of involving the elements of each sampled interval in the propagation process. The most intuitive way, is to treat these interval elements according to a uniform distribution; a choice that has been also adopted in previous research (Beer, Ferson et Kreinovich, 2013). We therefore consider the hypothesis that the interval elements follow a uniform distribution.

For a sampling level '$\alpha - cut$', $\alpha \in [0,1]$ (N levels), the proposed algorithm is as follows:

- Step 1: Determination of the interval boundaries resulting, from the '$\alpha - cut$' sampling process. These boundaries depend on the distribution shape and on the considered level '$\alpha - cut$'. Each interval element is assumed to follow a uniform distribution.
- Step 2: Calculation of reliability index on each '$\alpha - cut$' interval according to the approach proposed by Gagnon *et al.* (Gagnon et al., 2013).
- Step 3: Determination of the distribution associated with the reliability index and its associated statistical properties.

These steps are necessary for the reliability calculation, which requires that variables follow a probability distribution. Without this adaptation, the reliability calculation couldn't be performed with the current reliability model and its modification should be considered.

With this algorithm we extend the reliability calculation, according Hasofer-Lind approach, to variables resulting from imprecise probability theory. This approach allows reliability index results to be represented under a distribution shape, based on a $\alpha - cut$ sampling level (Figure 2.2). This approach requires some calculation steps, but in turn it allows us to handle directly the 'epistemic uncertainty' without any modification of the original information's formulation. If we are interested in having a single value, then we can aggregate all the results, by adopting for example a $\alpha - cut$ weighted approach. Nonetheless, distributions belonging to the possibility theory are generally bounded distributions and require corrections according to the approaches proposed in Section 2.3.4.

2.3.5.2 Evidence distributions processing

The expert opinions formulated according to the evidence theory are often presented as events with their belief degrees (discrete representation)(Ayyub, 2001). However, the current reliability model requires a continuous distribution. This forces us to add a hypothesis on this distribution in order to guarantee its continuity; otherwise it is necessary to solve the problem with another reliability method. An intuitive way is to assume that the variation between two consecutive event belief assignments is linear, which allows us to determine a pseudo-continuous cumulative density function for the evidence distribution (Figure 2.6).The distributions belonging to the evidence theory are generally bounded distributions and require corrections according to the approaches proposed in Section 2.3.4.

2.3.6 Results and discussion

This section discusses the reliability index results for the different proposed theories: uniform, Gumbel, triangular, evidence and possibility (Tableau 2.1). The theories characteristics and support domains are elaborated according to the quantile approach (Section 2.3.3), based on a given reference theory (Gumbel). Tableau 2.2 summarizes characteristics of the different theories for case G05. According to the quantile approach, the upper and lower bounds related to each theory support domain are slightly different from one distribution to another, much like their respective modes, which are not necessarily at the same location. This leads to a slight disparity in the reliability results obtained from these theories (Figure 2.8). It should also be noted that the belief function shape (in possibility and evidence theories) affects the calculation of theory support domain bounds, which results also in a slight reliability calculation disparity. If the theories support domains were identical for all case matrix theories, the reliability indices would be higher, compared to the approach based on quantile method.

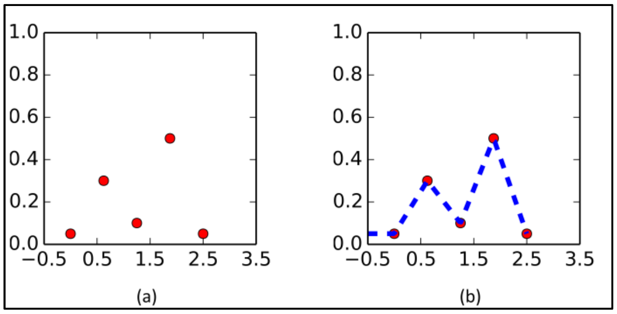


Figure 2.6 (a) Example of a belief function expressed as an evidence distribution and (b) example of a continuous belief function for the example in (a)

Tableau 2.2 Distribution characteristics for defect size and loading stress – G05

	Representation format	**Defect size** a – [mm]	**Loading stress range** $\Delta\sigma$ – [MPa]
Uniform	$[B_{low},\ B_{up}]$	$a \sim F_U(0.78, 3.54)$	$\Delta\sigma \sim F_U(24.55, 29.98)$
Gumbel	(Localisation=μ, Scale factor=β)	$a \sim F_G(1.52{,}0.514)$	$\Delta\sigma \sim F_G(21.01, 1.01)$
Triangular	$[B_{low}, mode, B_{up}]$	$a \sim F_T(0.53, 1.74, 3.82$	$\Delta\sigma \sim F_T(24.07, 26.44{,}30.5$
Evidence	$[E_1,\ldots, E_n]$ $[m_1,\ldots, m_n]$	$a \sim$ [0.78,....,3.54] [0.02,...,0.9,...0.01]	$\Delta\sigma \sim$ [24.07,.......,30.53] [0.02,...,0.9,...0.01]
Possibility	$[B_{low}, mode, B_{up}]$	$a \sim$ [0.65, 1.54, 2.84]	$\Delta\sigma \sim$ [24.49, 25.39, 27.85]

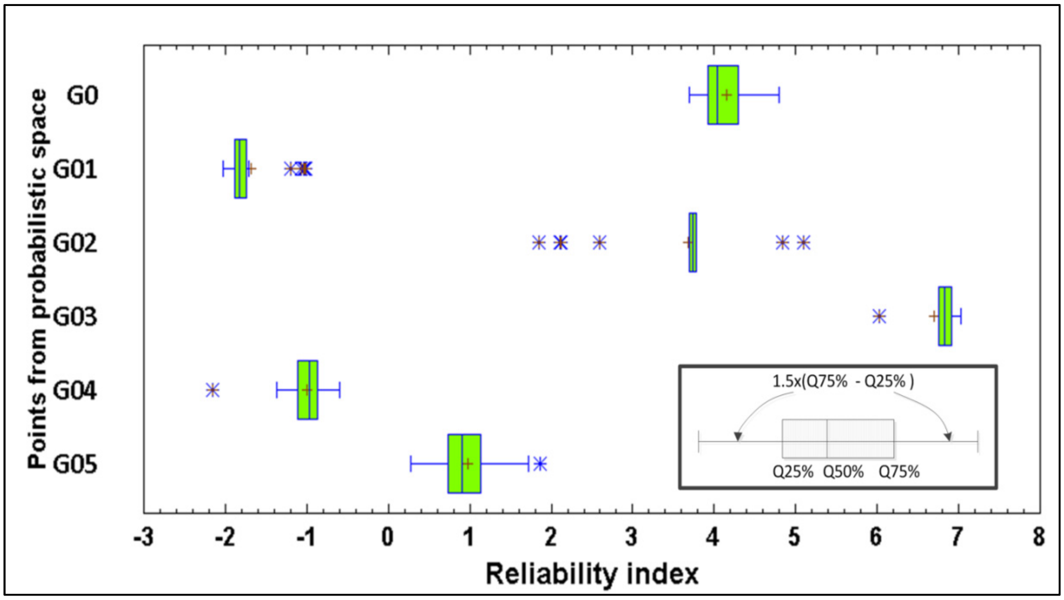


Figure 2.7 Reliability indices calculated according to different theories combinations (Tableau 2.1) for defect size and loading stress for G0, G01, G02, G03, G04 and G05 (Figure 2.4)

The propagation model complexity, the way that information has been processed and the weight of input variables in the model, are other factors which can explain disparity between theories. In fact more the system model is complicated, the more the resulting distribution shape diverges from their original variable distributions, which leads to higher disparity between theories. Hahn and Shapiro provide examples of some distribution combinations (Hahn et Shapiro, 1968). Also the gap between theories increases when one variable has a higher weight in the model, than the others variables. We can conclude that, for the studied reliability model, it is not the choice of the theory which affects the results, but rather the data formulation and the models chosen for their propagation.

Comparison between theories show that the uniform distribution is more conservative (lower reliability index) than the average of other distributions and the theory providing the highest reliability index for G05 is the evidence theory due to its distribution shape (see Figure 2.8). The reliability index mean value, resulting from the different theories for G05 (without uniform distribution), is 1.15, which is closer to the reference value for this case 1.18.

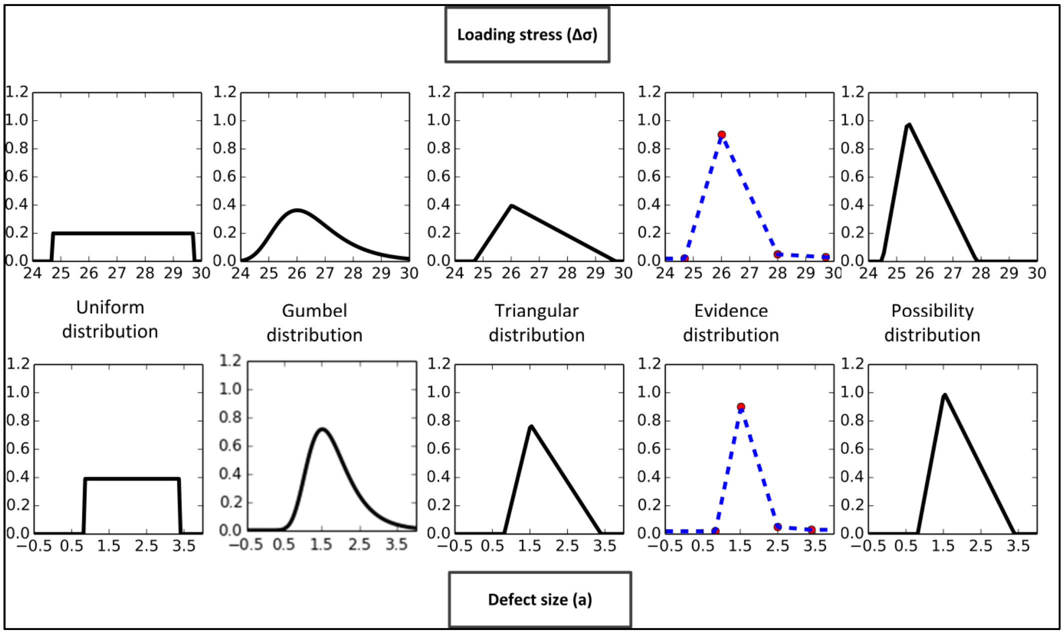


Figure 2.8 Distributions used in the case matrix

Reliability calculations resulting from triangular, evidence and possibility theories seem closer to the results provided by the Gumbel distribution, which is mainly explained by the proximity of their 'shapes' (Figure 2.8). Also we observed that for some operating points in the probabilistic space (Figure 2.4), that different theories provide similar results (Figure 2.7).

Furthermore, results showed that the lower reliability index value corresponds to the case where at least one input variable (i.e. defect size or loading stress) follows a probability distribution; a finding which support results advanced by Dubois et al. (Dubois et al., 2004). In fact for possibility theory, Dubois *et al.* have shown that the Necessity parameter corresponds to the lower probability bound and that the Possibility parameter corresponds to the higher probability bound (Dubois et al., 2004).

2.4 Conclusion

In this paper, we considered a fatigue reliability model based on the FORM approach and governed by a nonlinear limit state represented by the Kitagawa-Takahashi diagram.

Effective reliability assessment requires appropriate data, which in some cases need to be provided by an expert in order to fill a lack of information. These expert opinions could be expressed according to a probabilistic approach or according to imprecise probabilities. In FORM approach, input variables should follow a probability distribution, hence the proposed approach based on 'α-cut' sampling (Section 2.3.5.1) which allows the extension of the reliability calculation according to Hasofer-Lind methodology to variables derived from imprecise probabilities.

The comparison between the suggested theories (probabilistic theories and imprecise probability theories) was based on the quantile approach and on the reliability model proposed by Gagnon *et al.* (Gagnon et al., 2013). From these results, it was observed that the studied theories provide practically comparable results and the mixture of theories has no effect on reliability calculation, since they have equivalent support domains. Thus, for the studied reliability model, it is not the choice of the theory which affects the reliability calculation, but rather the data formulation and the models chosen for their propagation.

During case simulations, we faced some limitations associated to the FORM approach, when dealing with bounded distributions (e.g. triangular, possibility, uniform). In fact, the adoption of bounded distributions in the reliability calculation based on FORM approach affects the result sensitivity and the convergence of the calculation. To solve this, we suggested the approach described in Section 2.3.4 which mimics unbounded distributions at their tails. The other alternative consists of having at least one of the model input variables (defect size or loading stress), expressed by an unbounded distribution and preferably the variable with the largest variance and the highest weight in the reliability model. Also it was verified (by simulations) that this choice could be an acceptable condition (depending on the variable variances) to avoid bounded distribution limitations.

Sampling techniques, complexity of the propagation model, and the chosen membership function, are some factors which contribute in disparities between theories. This suggests putting more effort on the development of the propagation models as well as the elicitation and aggregation process. Therefore, for the studied reliability model, we should favor expert opinions expressed under an unbounded distribution.

The authors believe that the approaches proposed in this paper, can be applied in other areas where uncertain data are manipulated and modeled according to imprecise theories.

2.5 Acknowledgment

The authors would like to thank MITACS-Canada and Institut de recherche d'Hydro-Québec (IREQ) for their financial support of the project.

CHAPITRE 3

A COMPARISON BETWEEN AGGREGATION BEFORE AND AFTER PROPAGATION BASED ON A RELIABILITY MODEL

Mounia Berdai[1], Antoine Tahan[1], Martin Gagnon[2]

[1]École de technologie supérieure, 1100, rue Notre-Dame Ouest, Montréal (Québec) H3C 1K3 Canada
[2]IREQ, 1800 Boulevard Lionel-Boulet, Varennes (Québec) J3X 1S1, Canada

This chapter has been submitted for publication to the "Information Sciences" Journal (June 2017).

3.1 Abstract

Experts are often solicited to provide their opinions on systems unavailable inputs. In certain cases we can have several opinions for each input. In such situation, we ask ourselves what is the best way to combine these opinions? We can either combine them, for each input, before their propagation into the system model, or combine them after the propagation of each opinions combination separately. The purpose of this paper is to explore the differences between these two aggregation modes. For our reliability model, outcomes show that the location of the operating point in the probabilistic space and the divergence (gap) between expert opinions are the main factors explaining this difference. In this paper, we propose the Divergence Metric δ to measure the divergence between experts' opinions and we suggest the use of the 'Cumulative Distribution Averaging' as an aggregation rule. This rule seems suitable for probabilistic and non-probabilistic opinions and it avoids limitations encountered with expert opinions which are expressed according to bounded distributions.

Keywords : Aggregation, Divergence Metric, Area Metric, Cumulative Distribution Averaging, Reliability, Bounded Distributions.

3.2 Introduction

Expert opinions are often solicited in areas where there are neither clear standards nor well-developed theories. An example of this situation is the prediction of poorly understood phenomena. In such cases, experts can estimate the epistemic information by using the classical theory of probability (Helton et Johnson, 2011) or by using imprecise probabilities (Dubois et Prade, 1998). This last kind of theories seems more appropriate for the presentation of epistemic uncertainty (Canfield, 2006; Helton et Johnson, 2011; Salehghaffari et Rais-Rohani, 2013). For example, evidence theory allows mutual effects between events and the aggregation of data resulting from different uncertainty modelling theories, to be taken into account (Dubois et al., 2004; Klir et Smith, 2001). The effectiveness of the imprecise theories depends mainly on the choice of membership functions, the selection of the appropriate experts, and also on their awareness about the chosen theories (Beer, Ferson et Kreinovich, 2013; Oberkampf et al., 2004).

In the case of multi-variable systems, this diversification of theories can result in different expert opinions modelling. To be able to assess the system behaviour based on its propagation model, we need to combine the available experts opinions for each input in order to have a representative value. To achieve this purpose, the most intuitive method is to combine all experts opinions associated to each input into a single value, before propagating all system inputs into the system model (Figure 3.1- a). By following this approach, we can miss the individual expert's 'signature' characterizing each opinion. Alternately, for all system inputs we can consider the different possible combinations of their respective experts opinions, and then propagate each combination separately in the system model. In this case final outcome is achieved by combining all obtained results from different propagations (Figure 3.1- b). For other systems, experts can directly provide a final expected output; as shown in Figure 3.1-c. This situation may refer to two scenarios: a configuration where experts provide their opinions directly on the model output, or a configuration where the studied system has no propagation model. In such case, the question can be summarized by a simple aggregation problem.

The difference between the two aggregation modes and its effect on the final results are still unknown, especially for a nonlinear propagation model. Consequently, we raise the following question: what is the optimal way to merge these opinions? Merge them at the model input and then propagate the obtained data, or propagating them separately into the propagation model and then merge all obtained data at the end? In our knowledge, the difference between aggregation before and after propagation has not been directly addressed in the literature, but it was only highlighted by Cooke (Cooke, 2004), who reported that the difference between the two aggregation modes could be significant.

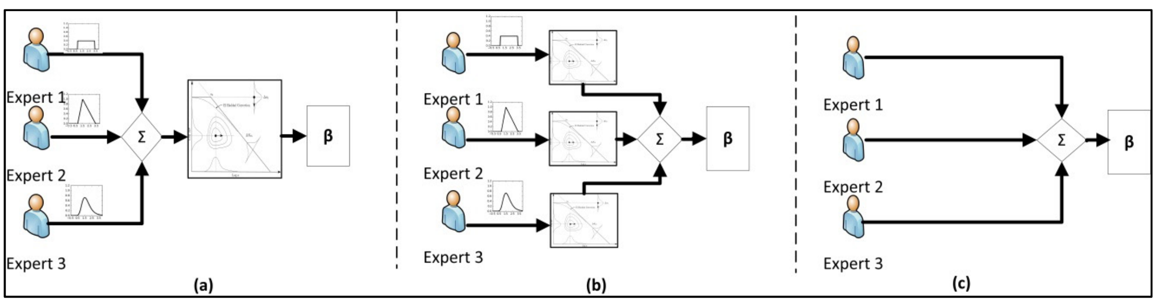


Figure 3.1 Different configurations of expert opinions fusion: (a) Fusion before propagation into the system model-(b) Fusion after propagation into the system model- (c) Experts directly provide model output

The purpose of this paper is to explore parameters explaining the possible differences that could be found between aggregation before propagation and after propagation for a given model. In our study this propagation model is represented by the fatigue reliability model proposed by (Gagnon et *al.*, 2013) and described in Section 3.4.1

The paper is structured as follows: Section 3.3 presents the methodology and the used assumptions. In this section, we discuss the Cumulative Distribution Averaging (CDA) concept and the proposed metric to quantify the gaps between expert opinions. In Section

3.4, we discuss the particularities of the study case, the proposed case matrix and the obtained results according to the methodology described in Section 3.3. Finally, Section 3.5 summarizes the conclusions of the study.

3.3 Methodology

The difference between aggregation before propagation and after propagation could be explained by many factors such as: the used aggregation rule, the gap between expert opinions, the propagation model, the choosing theories for modelling expert opinions and obviously the number of expert. As shown in Figure 3.1-a and Figure 3.1-b, some systems can have several experts available to elicit the same variable and in some cases the number of experts available for each system input could differ from one input to another, since each system input could require knowledge different from the other input. In this sense, if an unbalanced number of experts is considered, the difference between the two aggregation modes might also be affected.

In the current study, we use as a propagation model the reliability model described in section 3.4.1 with the following assumptions:

- A number of 3 experts is considered for each system input.
- Experts can express their opinions according to 3 probability distributions: Gumbel (G), triangular (T) and uniform (U) distributions.

Later in this section we present the concept of the aggregation rule based on the CDA and we discuss its relevance for the studied reliability model.

Finally as expert opinion divergence seem having an important effect on the difference between the two aggregation modes, we suggest a Divergence Metric (DM) to quantify this divergence between experts. The comparison between the two aggregation modes will be performed through the comparison of their respective reliability indices obtained according to the steps shown in Figure 3.2.

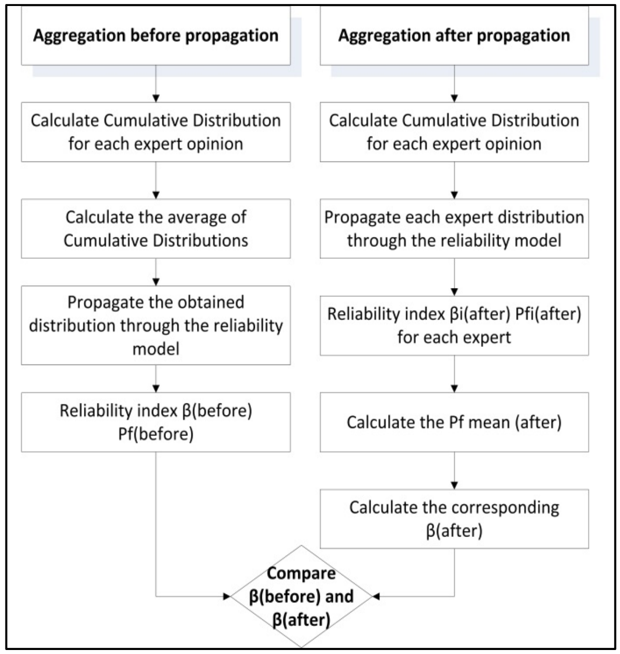


Figure 3.2 Approach followed in the comparison between the two aggregations modes: aggregation before propagation and aggregation after propagation

3.3.1 Cumulative Distribution Averaging (CDA)

The aggregation of the expert opinions is not an easy task because it involves both behavioural and mathematical aspects. Mathematical aggregation usually refers to approaches like: weighted averages, Dempster's rule or the Shannon (Ayyub, 2001; Lyu et al., 2010). These aggregation rules manage data differently, which can result in different findings. For example, in cases with fewer data, the results of the weighted average, geometric average, and median may differ from one approach to another (Meyer et Booker, 2001). Moreover, aggregation rules like the arithmetic mean, give more weight to the opinions of the majority, which can mask possible divergences or contradictions between experts, and may result in an empty intersection between opinions (Ha-Duong, 2008). In this case, the aggregation method based on the intersection of opinions cannot be applied. For these situations, Smets has proposed the TBM « *Transfer Belief model* » theory as a solution to this problem (Smets, 2007).

The choice of the aggregation rule also depends on the domain and the purpose of the study. For example, for some decision making processes, the aggregation can be performed according to a consensus, consensus ranking (Ben-Arieh et Chen, 2006), or according to a classification criteria (Meyer et Booker, 2001; Yuhua et Datao, 2005). In some studies, the aggregation has been performed according to preferential relationship (Wang et Fan, 2007), or it has been accomplished according to a specific algorithm (O'Hagan, 1988). In other situations, the aggregation was perceived as an optimization problem (Chen, Wang et Lu, 2011; Xu, 2004) involving specific aggregation operators such as the S-OWA, OR-LIKE operator (Yager et Filev, 1994), linguistic operator OWA (Yager, 1988), or order operator WGA(Xu et Da, 2003).

In the current study, we use a probabilistic model (Section 3.4.1) as a propagation model, which requires that its inputs follow probability distributions. Therefore, we adopt an aggregation approach based on arithmetic mean of probabilities, defined for n experts E_i, $i \in [1, n]$ as follows:

$$F_{CDA}(x_0) = \sum_{i=1}^{n} \omega_i F_i(x_0) \tag{3.1}$$

Where F_i represents the distribution used by the expert E_i to model his opinion, which could also be a pseudo-cumulative density distribution if the opinion is expressed according to a non- probabilistic distribution. F_{CDA} is the Cumulative Distribution Averaging (CDA) of the n distributions and ω_i represents the weight given to each distribution F_i. In our study, all experts have the same weight $\omega_i = n^{-1}$. Figure 3.3 shows an example of the resulting CDA of four different distributions.

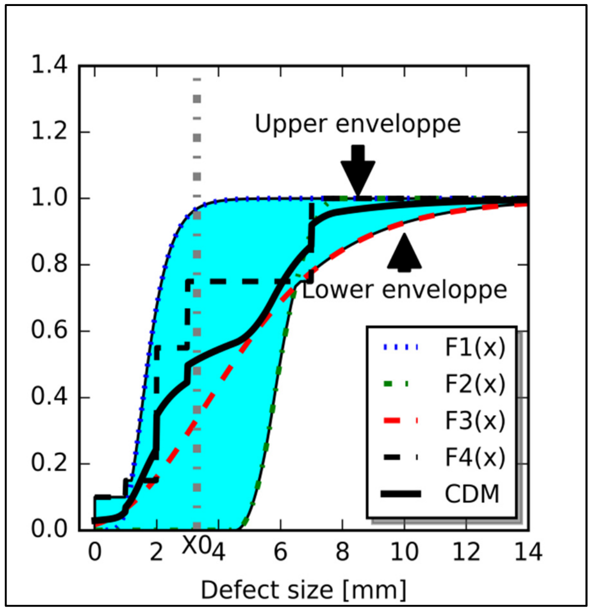


Figure 3.3 Example of CDA resulting from four different distributions

The proposed CDA aggregation rule can be used in any system models, since expert opinions are not always expressed as a single value, but they can follow a probabilistic distribution or imprecise probabilities (Dubois et Prade, 1998; Hsu et Chen, 1996). An aggregation rule similar to the proposed CDA has been proposed and used by Cooke *et al.* (Cooke et Goossens, 2008). However, in their research, weights ω_i were provided by analysts or were derived from experts' calibration and information scores.

In the case of an aggregation after propagation (the scenario shown in Figure 3.1 - b), each expert provides an opinion which is propagated through the reliability model independently. In such case, the combination of all outputs is performed according to the 'classical' arithmetic mean of all obtained individual probabilities and then the reliability index is calculated according to the obtained probability value (see Figure 3.2).

3.3.2 Gap between experts opinions assessment

In this study, we have explored how gap (divergence) between expert opinions influences the difference between the two aggregation modes before and after propagation (Figure 3.9); thus the need for an objective metric to quantify this difference. This can be performed through statistical tests such as the *t*-Student test. However, such tests do not provide a quantitative and continuous measurement of the agreement between experts, but rather just reject, or not, the null hypothesis.

In the decision-making domain, the agreement among experts has often been measured by using the proximity measure or the ranking approach, in order to assess the degree of consensus between experts opinions (Ben-Arieh et Chen, 2006; Herrera-Viedma, Herrera et Chiclana, 2002). In our study, such strategies can introduce subjectivities by adding bias to the results and thus make results analysis difficult.

In our case, we believe that the concept of Area Metric (AM) proposed by Ferson et *al.* can be a suitable alternative to quantify the agreement/disagreement between distributions. This concept is defined as follows (Ferson, Oberkampf et Ginzburg, 2008):

$$d(F_{xi}^{e}, F_{xi}^{m}) = \int_{-\infty}^{+\infty} |F_{xi}^{e}(x) - F_{xi}^{m}(x)| dx \tag{3.2}$$

Where F^e, F^m represent the cumulative density functions of the compared distributions.

This method uses the area between distributions to quantify the agreement/disagreement between them. Therefore, a large area reflects a large difference between the two distributions and a narrow area reflects the concordance between the two distributions. One distinctive advantage of this metric is its capability to use any kind of distribution and to measure disagreements which other methods using the lower-order moments like the mean and variance cannot address (Ferson, Oberkampf et Ginzburg, 2008).

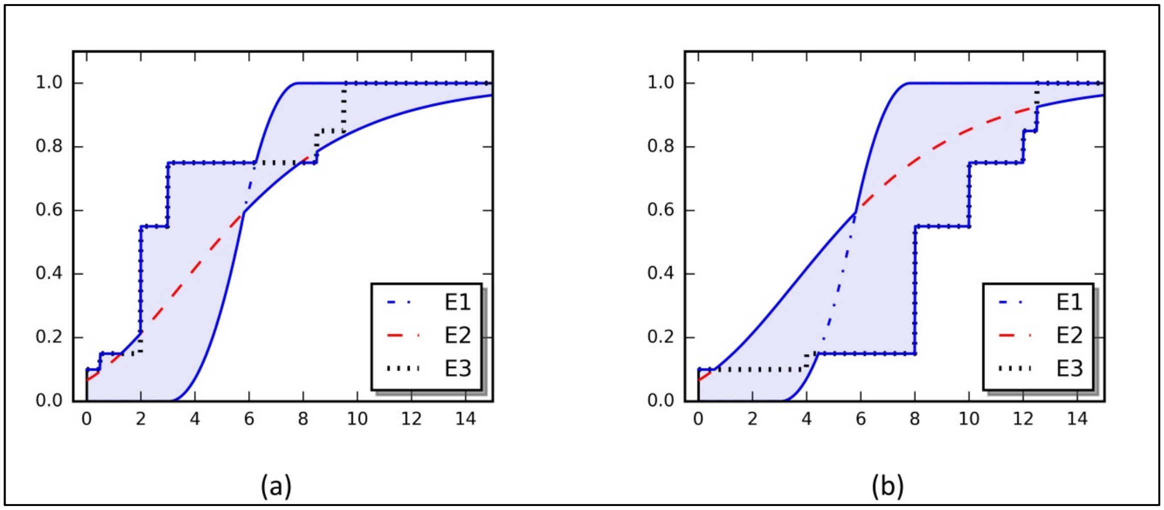


Figure 3.4 Example of *Area Metric* measure, based on 3 different distributions

As shown in Figure 3.4 ((a) and (b)) the AM changes according to the distribution shapes and to the differences between them; which accurately reflects the agreement between the distributions. Therefore, this metric can be considered as an objective way to describe and quantify the difference between any distributions and thus between any expert opinions (ASME, 2012; Liu et al., 2011).

For our purposes we define the DM $\delta(O_i, O_j)$ based on the AM concept, to estimate the divergence between expert E_i and expert E_j, knowing that the expert E_i provides the opinion $O_i \sim F_i$ formulated according to the distribution F_i and the expert E_j provides the opinion $O_j \sim F_j$ expressed according to the distribution F_j. This DM is defined as follows:

$$\delta(O_i, O_j) = \int_D \left| F_{O_i}(x) - F_{O_j}(x) \right| dx \qquad (3.3)$$

Where D represents the common support domain for both expert opinions.

From this relationship, we deduce that for $\forall\, i, j, k \in N$:

- $\delta(O_i, O_j) = \delta(O_j, O_i)$
- $\delta(O_i, O_i) = 0$
- $\delta(O_i, O_j) \leq \delta(O_i, O_k) + \delta(O_k, O_j)$

In Section 3 some practical examples of the use of this divergence metric are presented.

3.4 Study case

In this section we present a description of the fatigue reliability model which represents the propagation model. Then we describe the different cases considered in the study. Finally, in section 3.4.3 we provide obtained results according to the approach described in section 3.3.

3.4.1 Reliability model description

The studied reliability model is based on a probabilistic approach characterizing the fatigue problem of hydroelectric turbine blades (Gagnon, 2013). This model uses the First Order Reliability Methods (FORM) and the Hasofer-Lind approach to calculate the reliability index β_{HL}. The failure probability P_f is deduced from the calculated reliability index β_{HL}, according to the following equation: $P_f \approx 1 - \varphi(\beta_{HL}) = \varphi(-\beta_{HL})$, where φ is the standard Gaussian cumulative function. The limit state $g(x)$ used in this reliability model is the threshold determined by the Kitagawa-Takahashi diagram (Figure 3.5). This limit is function of material mechanical properties, fabrication process and residual stress resulting from the welding process (Thibault, Gagnon et Godin, 2014). The comparison between the aggregation before and after propagation is performed through the reliability model as shown in Figure 3.2.

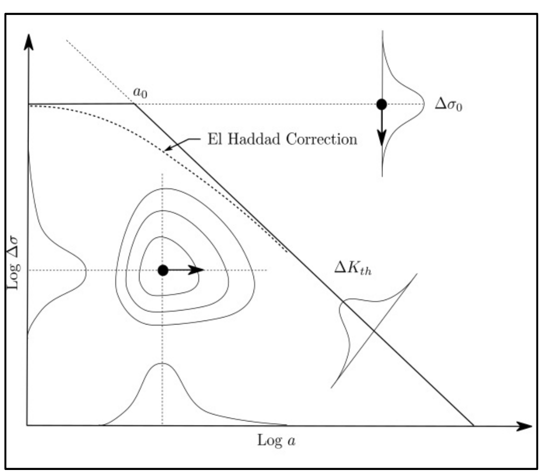


Figure 3.5 Kitagawa-Takahashi diagram

For this reliability model, we can distinguish two main regions: a safe region which is under the limit state function $g(x)$ and an unsafe region which is above the limit state function (Figure 3.5 and Figure 3.6). The probability of failure and the reliability index depend on the location of the operating point, which is defined by the loading stress $\Delta\sigma$ and the defect size a (Tableau 3.1). For example, operating points in the safe region, far from the limit state $g(x)$ (like G1 in Figure 3.6) have a very low failure probability ($P_f \approx 10^{-6}$) and points in the unsafe region, far from the limit state (like G2 in Figure 3.6) have a high failure probability ($P_f > 0.5$). This means that all points closer to G1 and G2 will have similar reliability indices. Thus, an analysis solely based on these two extreme operating points would not be representative to explain the difference between the two aggregation modes. Therefore, additional operating points G3 and G4 have also been studied, as shown in Figure 3.6. These additional operating points allow more coverage of the probabilistic space. This set of operating points (G1, G2, G3 and G4) respectively describes the progression from the safe to the unsafe region and reflect the change in the failure probability and the reliability index when we move toward and away from the limit state. Characteristics of these points are presented in Tableau 3.1.

Tableau 3.1 Distribution characteristics for defect size and loading stress, for G1, G2, G3 and G4. They are expressed according to Gumbel distribution F_G (Location= μ, Scale= β)

	Defect size $a - [mm]$	**Loading stress range** $\Delta\sigma$ – [MPa]
G1	$a \sim F_G(1.0,0.5)$	$\Delta\sigma \sim F_G(15.0,1.0)$
G2	$a \sim F_G(2.0,0.5)$	$\Delta\sigma \sim F_G(34.0,1.0)$
G3	$a \sim F_G(6.0,0.5)$	$\Delta\sigma \sim F_G(15.0,1.0)$
G4	$a \sim F_G(1.5,0.5)$	$\Delta\sigma \sim F_G(34.0,1.0)$

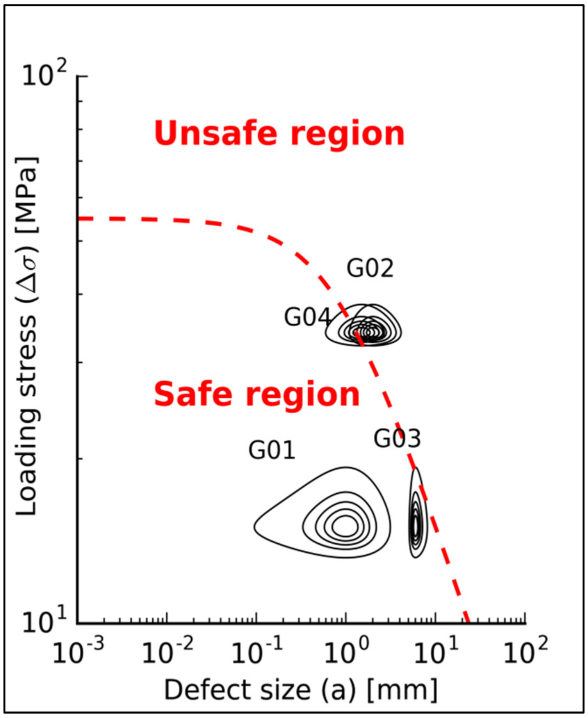


Figure 3.6 Operating points locations in the probabilistic space

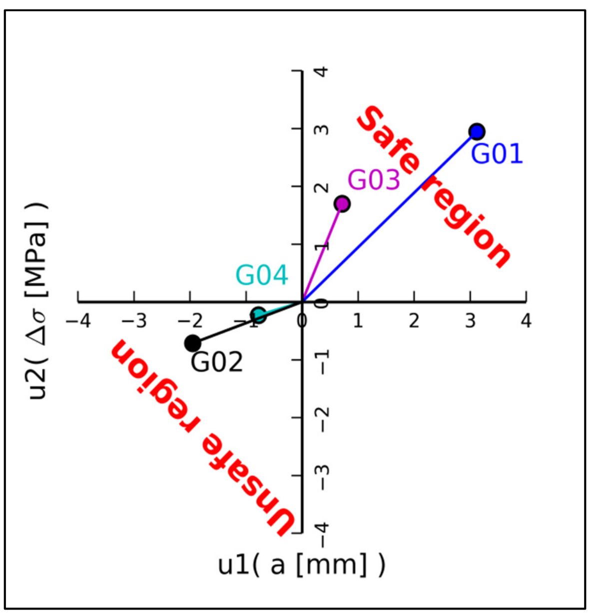


Figure 3.7 Reliability index of the operating points from Figure 3.6, in the iso-probabilistic space

3.4.2 Studied expert opinions cases

In order to explore parameters which influence the differences between the two aggregation modes, we consider 3 experts who are assumed to express their opinions according to 3 probability distributions: Gumbel (G), triangular (T) and uniform (U) distributions (Figure 3.8), as discussed in Section 3.3. By considering all possible combinations we obtain 27 cases as shown in Figure 3.8. By adding 4 possible gaps between experts (Figure 3.9) and 4 operating points in the probabilistic space (Figure 3.6), we end up with 27×4×4 cases to be evaluated.

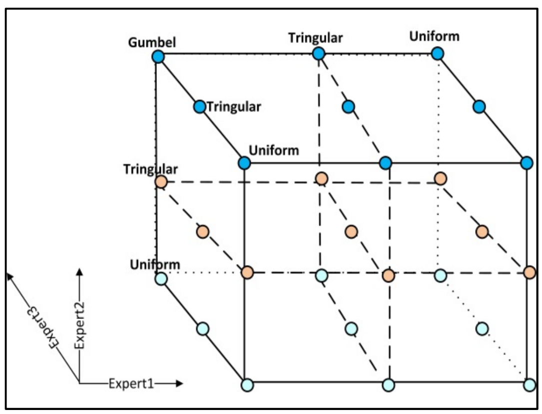


Figure 3.8 Matrix of the 27(3^3) sensitivity cases: 3 experts to elicit the defect size according to 3 distributions: Gumbel (G), triangular (T) and uniform (U)

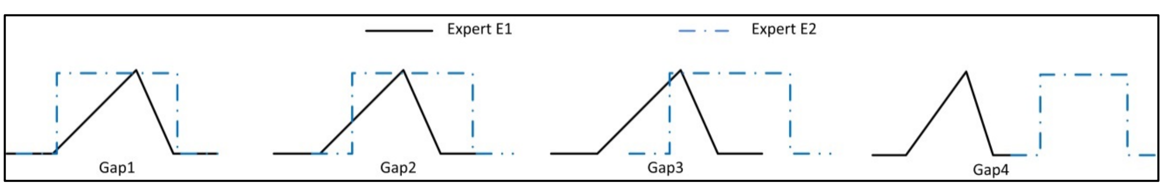


Figure 3.9 Illustration of the 4 considered gaps between 2 experts. In this example the opinion of expert E_1 follows a triangular distribution and the opinion of expert E_2 follows a uniform distribution

In Figure 3.9, ‘Gap1’ indicates that there is no gap between experts, ‘Gap2’ refers to a slight gap, ‘Gap3’ refers to a higher gap and ‘Gap4’ refers to distinct opinions of the two experts. These gaps generate differences among expert opinions varying from 5% to 50%, according to the DM δ suggested in section 3.3.2.

3.4.3 Results

Based on the DM δ proposed in section 3.3.2, we quantified the divergence between experts, as shown in Figure 3.10. In this figure, the DM was calculated between the 3 experts at the operating point G1.

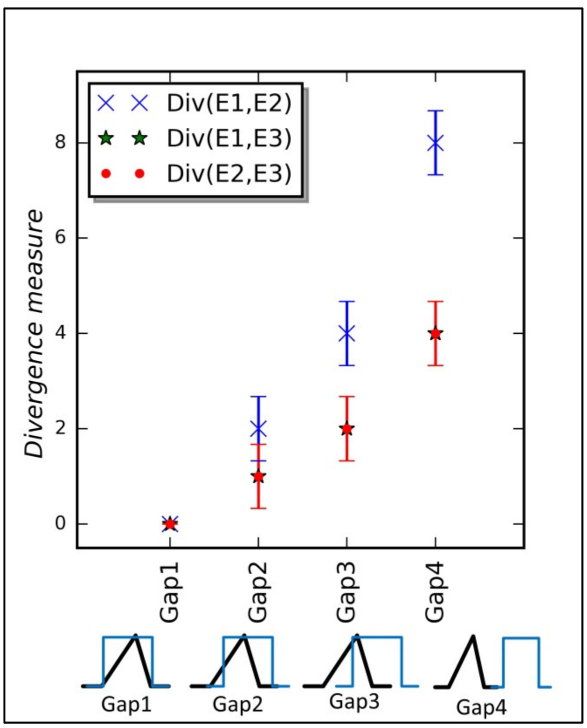
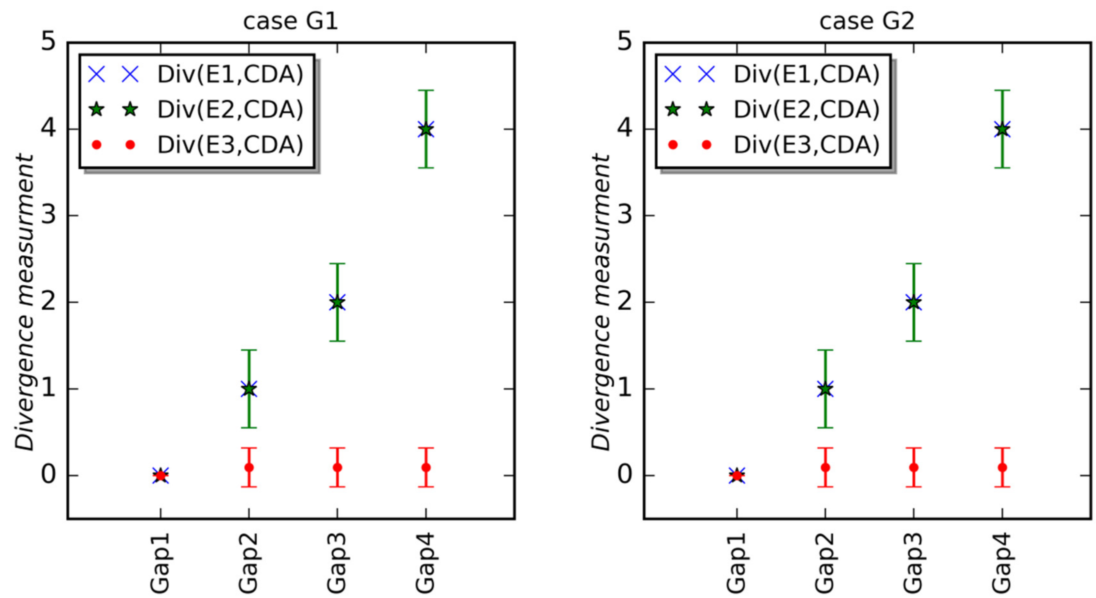


Figure 3.10 Divergence Metric (DM) between opinions from the 27 possible combinations of the 3 expert opinions, corresponding to the operating point G1

Results from Figure 3.11 are based on the DM coefficient and reflect the consistency of the approach adopted in this study for the processing of the four considered operating points.

Also from this figure (Figure 3.11) we can observe that opinions of expert E3 seem to be the closest ones to the group average, for all operating points and for all considered gaps. Thus the DM can be considered as a tool to 'identify' and to 'to classify' experts, according to their proximity to the 'group average'. This classification will help later in understanding factors influencing the expert's divergence from the group average, when consensus (for example) is adopted as an aggregation rule. The DM can then constitute an objective measurement of the consensus between experts and a good alternative to what has been suggested in other studies (Ben-Arieh et Chen, 2006) .



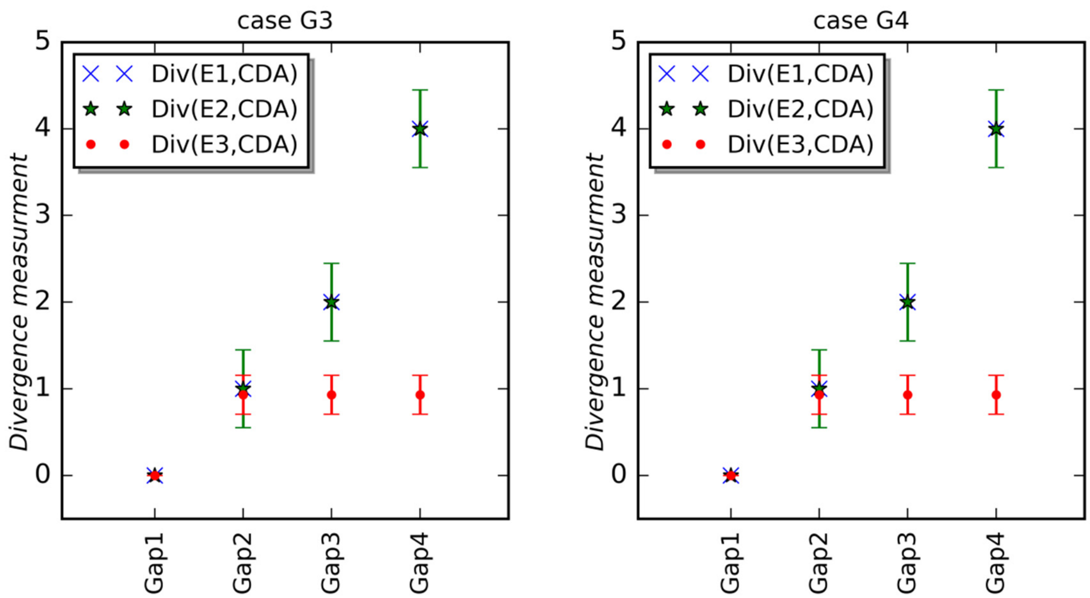


Figure 3.11 Divergence measurement between experts and the CDA of the group, for operating points: G1, G2, G3 and G4 considering the 27 possible combinations of the 3 expert opinions

When the operating point is in the safe region far from the limit state (case G1 in Figure 3.6) the reliability index is higher, and consequently the failure probability is very low ($P_f \approx 10^{-6}$). In these conditions, the difference between aggregation before and after propagation is higher which is mainly explained by the model sensitivity and not necessarily by the difference between the two aggregation modes. On the other hand, when the operating point is in the unsafe region (case of G2), the failure probability is higher ($P_f \geq 0.5$) and once we start to move away from the limit state, the failure probability at any point in the unsafe region becomes closer to 1 ($P_f \approx 1.0$). In such situation, all points in the unsafe region have the same reliability index β.

For the operating point G3, once we start to move towards the limit state, the failure probability becomes significant and cannot be neglected. Also, for this location, we notice that, when the gap between experts is high, the difference between the two aggregation modes becomes important (Figure 3.12) resulting in more conservative values for the aggregation before propagation than the aggregation after propagation. Alternately, in the unsafe region (G2 and G4 locations), the difference between the two aggregation modes seems insignificant and not very sensitive to the gap between experts, as shown in Figure 3.12.

From Figure 3.12 we notice that if we are in the safe region, the effect of gaps between experts becomes significant on the difference between the two aggregation modes. In fact, in the safe region and for a large gaps (e.g. Gap4), as soon as we move towards the limit state, the resulting CDA shape changes. This shape depends on the chosen distribution for opinions modelling, the used propagation model, and the gap between the considered experts. Thus, in this specific region, the larger the gap between experts is, the more the resulting CDA is distorted, as shown in Figure 3.13 (case of (a), (b) and (c)).

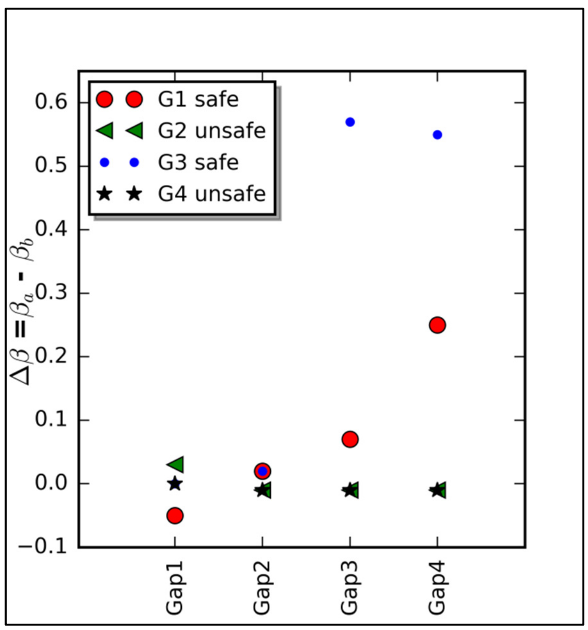


Figure 3.12 Differences of reliability indices β, between aggregation after and before propagation for each gap and each operating point. $\beta_b = \beta_{before}$ aggregation, $\beta_a = \beta_{after}$ aggregation

Tableau 3.2 Resulting reliability indices for the aggregation before and after propagation, for the four gaps and the four considered operating points in the probabilistic domain

	Gap1	**Gap2**	**Gap3**	**Gap4**

	$\boldsymbol{\beta_b}$	$\boldsymbol{\beta_a}$	$\boldsymbol{\beta_b}$	$\boldsymbol{\beta_a}$	$\boldsymbol{\beta_b}$	$\boldsymbol{\beta_a}$	$\boldsymbol{\beta_b}$	$\boldsymbol{\beta_a}$
G1	4,45	4,40	3,48	3,50	2,50	2,57	0,75	1,00
G2	-2,13	-2,10	-2,51	-2,51	-2,51	-2,51	-2,51	-2,51
G3	1,66	1,65	1,02	1,04	0,33	0,91	0,31	0,90
G4	-0,95	-0,95	-1,57	-1,58	-1,57	-1,58	-1,57	-1,58

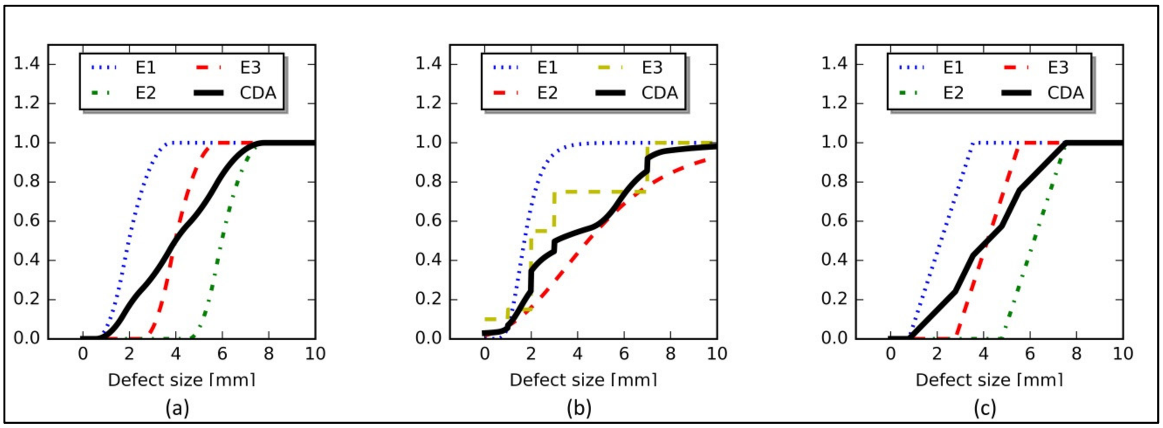


Figure 3.13 CDA of different expert opinions with different gaps and different distributions.

When we cross the limit state of the reliability model, the difference between the two aggregation modes changes its behaviour. In fact, in the unsafe region far from the limit state, almost all operating points have their failure probability closer to 1 ($P_f \approx 1.0$). Consequently, the difference between the two aggregation modes does not change. Contrary to the points in the ‘safe region’, the points in the ‘unsafe region’ showed that the aggregation before propagation is less conservative (slightly higher reliability index) than the aggregation after propagation, with an absolute error of 1%. Therefore we can state that difference between the two aggregation modes becomes significant only when we are in the safe region closer to the limit state.

3.5 Conclusion

We conducted this study to elucidate the difference between aggregation before and after propagation. Findings show that 'gaps' between expert opinions are identified as the main factors explaining the difference between the two aggregation modes. For our specific propagation model (fatigue reliability model section 3.4.1), the location of the operating point in the probabilistic space is also pointed as important factor influencing the difference between the two aggregation modes. This influence becomes significant when we are in the safe region and closer to the limit state of the reliability model. Otherwise, this difference is insignificant. Moreover, our results show that the aggregation before propagation is more conservative than aggregation after propagation for the studied reliability model which makes it appropriate for decision making process. Also, the choice of the aggregation before propagation will reduce the number of calculation steps by a ratio proportional to the square of the expert's number.

To assess gaps between experts we proposed a new metric δ (DM), based on the Area Metric concept. This metric measures the difference and divergence between expert opinions and can also be used as a tool to quantify the 'consensus degree' between experts, which will help in reaching a quick and objective agreement when consensus is required.

Moreover, we proposed the use of CDA as an aggregation rule, which is suitable for both probabilistic and non-probabilistic opinions. It also avoids limitations encountered with bounded distributions as shown in (Berdai, Tahan et Gagnon, 2016). In fact by choosing the aggregation before propagation and the CDA as an aggregation rule, we contribute in increasing the resulting distribution order moments, such as skewness and kurtosis, which lead to 'fat' tails, reducing the limitation of 'bounded distributions' and avoiding special processing required to adapt them to the reliability models using on the FORM approach.

3.6 Acknowledgments

The authors would like to thank MITACS-Canada and *Institut de recherche d'Hydro-Québec* (IREQ, Québec, Canada) for their financial support of the project.

CHAPITRE 4

REPRODUCTIBILITY INVESTIGATION OF ELICITATIONS TECHNIQUES IN RISK ASSESSMENT FOR HYDRAULIC TURBINES

Mounia Berdai[1], Antoine Tahan[1], Martin Gagnon[2]

[1]École de technologie supérieure, 1100, rue Notre-Dame Ouest, Montréal (Québec) H3C 1K3 Canada
[2]IREQ, 1800 Boulevard Lionel-Boulet, Varennes (Québec) J3X 1S1, Canada

This chapter has been submitted for publication in the *"Journal of risk research"*
(August 2017)

4.1 Abstract

Certain elicitation techniques exert some control on expert opinions by leading them to a consensus or to a specific choice. In the absence of such guidelines, experts rely on their own knowledge to formulate opinions. This can result in large dispersions and affects the decision maker judgement. In this situation, we wonder what the relevant elicitation techniques are and how we can help experts to express their knowledge. From literature revue, it is hard to decide if elicitation techniques are equivalent or not, which justifies the reproducibility analysis that we carry out in this paper. In this study, multiple experts have been involved in order to predict the defect size in hydraulic turbines, according to four proposed elicitation techniques. The comparison between these techniques was performed based on a suggested algorithm using the area metric concept. Our Findings show that elicitation techniques with 'support' tend to limit variations between experts and might be suitable only when prior knowledge on the expected elicited variable is available. Otherwise, we can end up with a distorted opinion of the elicited variable and an erroneous risk assessment.

Keywords: Decision making, elicitation technique, risk assessment, fatigue, reproducibility, reliability.

4.2 Introduction

Expert opinions are often solicited in areas where there are no clear standards, data or models, such as the prediction of poorly understood phenomena. In these cases, experts' contributions can increase the model's precision and facilitate decision-making in an optimal, cost-effective manner (Kuhnert, Martin et Griffiths, 2010). This could be obtained through a formal elicitation process according to series of specific questions (Ayyub, 2001; Clemen et Winkler, 1999). However, the lack of standards governing the elicitation process can lead to a variability in the resulting expert opinions due to the differences in their background and respective knowledge (Meyer et Booker, 2001). In certain cases, this difference generates disagreement or conflict between expert opinions which might influence the decision maker's opinion. However, 'guiding' experts with a suitable framework and elicitation technique can contribute in reducing these possible disparities at the risk of imposing some bias.

From the literature, we identified elicitation techniques such as: Triadic sorting , Free sorting, Direct sorting, and Ranking or Picking from an attribute list (Bech-Larsen et Nielsen, 1999). These combined with visual techniques or group interviews seem more suitable for cases where consensus is required (Breivik et Supphellen, 2003). In such situations, the consensus will reflect a high level of agreement and can mask the effect and particularities of certain elicitation techniques. In this sense, we notice that few studies compared and identified suitable elicitation techniques for specific domains. We can mention the study of Davis *et al.* who performed a comparative study applied to the software domain and demonstrated that elicitation techniques are not necessarily interchangeable (Davis *et al.*, 2006). In the environmental domain, O'Leary *et al.* have performed the comparison of three elicitation techniques and showed that opinions depend on the chosen elicitation technique (O'Leary *et al.*, 2009). To elicit consumer behaviour, Bech-Larsen *et al.* have used five elicitation techniques and they showed that the difference between elicitation techniques is not significant regarding predictive ability. As a results they recommended using the less expensive technique (Bech-Larsen et Nielsen, 1999). The same observation has been noted by Breivik *et al.* who concluded that there is no significant difference between elicitation techniques, but the elicitation method effects can be category-dependent (Breivik et Supphellen, 2003). From these studies, it is hard to decide whether elicitation techniques are interchangeable, especially if the studied domain or the assessment criteria are modified.

The purpose of this study is to analyse and assess elicitation techniques for the prediction of the likely defect size in hydraulic turbine runner blades. The defect size is one of the main inputs for the studied fatigue reliability model and it depends on different factors which make its modelling and prediction difficult. For such cases, experts might benefit from some support during the elicitation process, since the available empirical elicitation bases are very limited (Dieste et Juristo, 2011). This support should help experts to encode their knowledge without forcing them towards a particular choice.
Our objectives are to:

- Evaluate the reproducibility of the proposed elicitation techniques.
- Assess the impact and effectiveness of the suggested ‘supports’.
- Assess the elicitation techniques effectiveness for the proposed elicitation of the defect size.

The paper is structured as follows: Section 4.3 presents the methodology followed in this paper. It describes how the elicitation process was conducted and introduces the suggested elicitation techniques with the proposed algorithm for their assessment. Section 4.4 summarizes obtained results and Section 4.5 discusses these outcomes. Finally, Section 4.6 provides some conclusions.

4.3 Methodology

An accurate risk assessment based on the reliability model proposed by Gagnon *et al.* (Gagnon *et al.*, 2013), requires an up-to-date measurement of the likely defect size in hydraulic turbines blades. Often this information is hard to obtain due to the hydraulic turbine operational constraints, which suggest the use of expert opinions to fill this lack of information. For this reliability model no suitable elicitation technique has been developed or identified previously. Thus the purpose of the current paper aiming to compare some proposed elicitation techniques.

The methodology followed in this paper consists in defining:

- The elicitation process.
- The proposed elicitation techniques.
- The algorithm for their comparison.
- The leverage coefficient for expert opinions analysis.

4.3.1 Elicitation process

In this study, experts were asked to provide their best estimation of the maximum defect size in the hydraulic turbine runner blades. The defect size is one of the main inputs of the reliability model proposed by Gagnon *et al.* (Gagnon *et al.*, 2013), which calculates the reliability index β_{HL} (and the failure probability P_f) vis a vis its limit state $g(\boldsymbol{x})$ according to the provided inputs: the loading stress ($\Delta\sigma$ – [MPa]) and the defect size ($a - [mm]$). This limit state $g(x)$ represents the threshold determined by the Kitagawa-Takahashi diagram. It depends on the crack growth threshold $\Delta \mathrm{K_{th}}\left[\mathrm{MPa}\sqrt{\mathrm{m}}\right]$ and the fatigue limit $\Delta\sigma_0$ [MPa] (Figure 4.1).

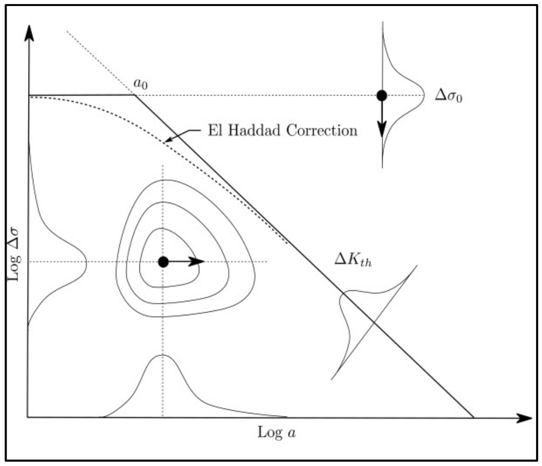


Figure 4.1 Kitagawa-Takahashi diagram

For a hydraulic turbine, typically the high stress areas are generally located at the welded junctions, connecting the blades to the runner crown and to the runner band. Figure 4.2 shows the simplified loading profile used in this study. It is composed of, *Low Cycle Fatigue* LCF stress cycles, σ_{LCF} and *High Cycle Fatigue* HCF stress cycles σ_{HCF}.

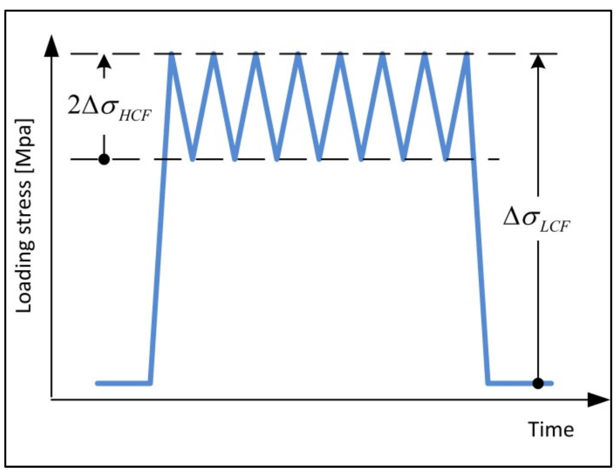


Figure 4.2 Simplified loading stress profile

In this study, 16 elicitation cases were proposed consisting in the combination of four elicitation techniques (A, B, C & D section 4.3.2) with other four operating conditions, as shown in Tableau 4.1.

Tableau 4.1 Operating conditions for the four elicitation techniques

	Case #1	**Case #2**	**Case #3**	**Case #4**
$\Delta\sigma_{LCF}$ [MPa]	100	150	100	150
$\Delta\sigma_{HCF}$ [MPa]	25	25	25	25
Number of service years [years]	10	10	30	30
Number of start-ups per day	1	1	1	1

In total, five experts were selected for this study, in order to assess the expected radius of a circular defect size $a[mm]$ as shown in Figure 4.3, for the 16 proposed elicitation cases, as shown in Tableau 4.1. The defect size (and its propagation) relies on many factors such as operating conditions and maintenance history. Consequently, experts involved in this process should have suitable knowledge of all these aspects in order to provide appropriate opinions for the defect size prediction. The standard BS7910 (BSI, 2005) was given as a reference to the experts if they judge it useful for the defect size prediction.

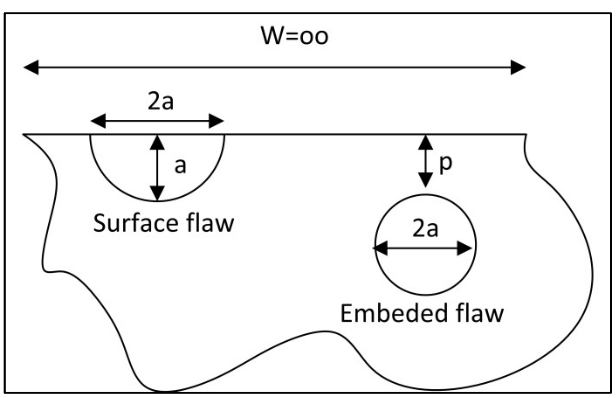


Figure 4.3 Surface flaw and embeded flaw

Two categories of experts have been considered: experts from the academic-research field and experts from the industrial field. All of them are working on subjects related to turbine runner blade fatigue reliability. After confirming their participation in the elicitation process, a support document describing the subject and the required information was sent to each expert. Then a kick-off meeting was organised to clarify the scope of the project and to collect their feedback on the process. This meeting was also an opportunity to study an example of the proposed elicitation forms and to show experts how they can encode their knowledge. We also discussed their concerns about the elicitation subject and the required data format. Following this meeting, experts received the appropriate elicitation form for each elicitation technique separately and were asked not to communicate among themselves during the elicitation process.

4.3.2 Elicitation techniques

The elicitation techniques in the literature can be classified under two categories: direct or indirect techniques. In direct techniques (like Probability Estimate, Probability Distribution and Bayesian Updating (Meyer et Booker, 2001; O'Hagan, 2012; Rocquigny, 2012), experts are able to articulate their judgment through direct questions. In contrast, indirect techniques assume that relevant information might not be easily accessible. To overcome this limitation, indirect techniques use approaches based on similarities and differences to model the required information (Hudlicka, 1996). Unfortunately, proximity estimation constitutes a relative measurement which is linked to a specific context (Cooke, 1994), therefore the adoption of these techniques should be used carefully; otherwise the expert may have difficulty understanding the purpose of the questions due to the lack of transparency which may influence answer quality (Hora, 1992). Indirect techniques are considered mostly for variables with limited data or with unavailable measurements which make their quantification by direct elicitation difficult (Cooke et Goossens, 1999). In this category, we find techniques as: Pairwise Comparisons, Ranking, Rating, Triadic Sorting, Direct Sorting, Picking from an attribute list, Clustering (Bech-Larsen et Nielsen, 1999; Chen et Pu, 2004; Meyer et Booker, 2001). It should be noted that indirect techniques are more time-consuming because they involve the analysis of a combination of concepts (Cooke, 1994). Moreover, some of these techniques impose a particular expression format to expert judgment, which risk generating bias in expert opinion (Cooke et Goossens, 2004). Montibeller *et al.* and

Meyer *et al.* presented a detailed review of various bias sources and some debiasing techniques (Meyer et Booker, 2001; Montibeller et Winterfeldt, 2015).

Faced with the risk of generating bias and in order to handle imperfections associated to expert knowledge, some studies advocate the use of direct techniques based on probability theory as much as possible (Meyer et Booker, 2001). However, if the data presents epistemic uncertainty, it might be better to choose an adapted framework to model such knowledge (Meyer et Booker, 2001). Frameworks like intervals theory or evidence theory allow, in such cases, a better representation of expert knowledge (Destercke, Dubois et Chojnacki, 2006; Fallet *et al.*, 2011). This has also been supported by Mannhart *et al*. who proposed the use of evidence theory and fuzzy logic for epistemic data modeling (Mannhart, Bilgic et Bertsche, 2007), while other authors have suggested adopting probability distributions combined with Bayesian updating for cases requiring multiple estimates (Meyer et Booker, 2001).

Also the choice of the suitable elicitation technique should take into account the studied domain, the nature of the expected information, the knowledge level of available experts and how this elicited information will be incorporated into the system model (Kuhnert, Martin et Griffiths, 2010). Furthermore, some studies highlight other aspect like the experts' mapping skills, time constraints and available funding (O'Leary *et al.*, 2009). Based on this discussion, we have suggested four elicitation techniques to attend experts in predicting the likely defect size in hydraulic turbines runner blades, for each operating conditions as shown in Tableau 4.1. The proposed techniques can be grouped under two main categories as presented in Figure 4.4: techniques A and B are based on a belief function, while techniques C and D are based on probability distributions. By comparing technique A with technique B, and technique C with technique D, we will measure the influence of the provided 'support' on the expert opinion. This comparison will also help us to refine the key parameters chosen to support the experts.

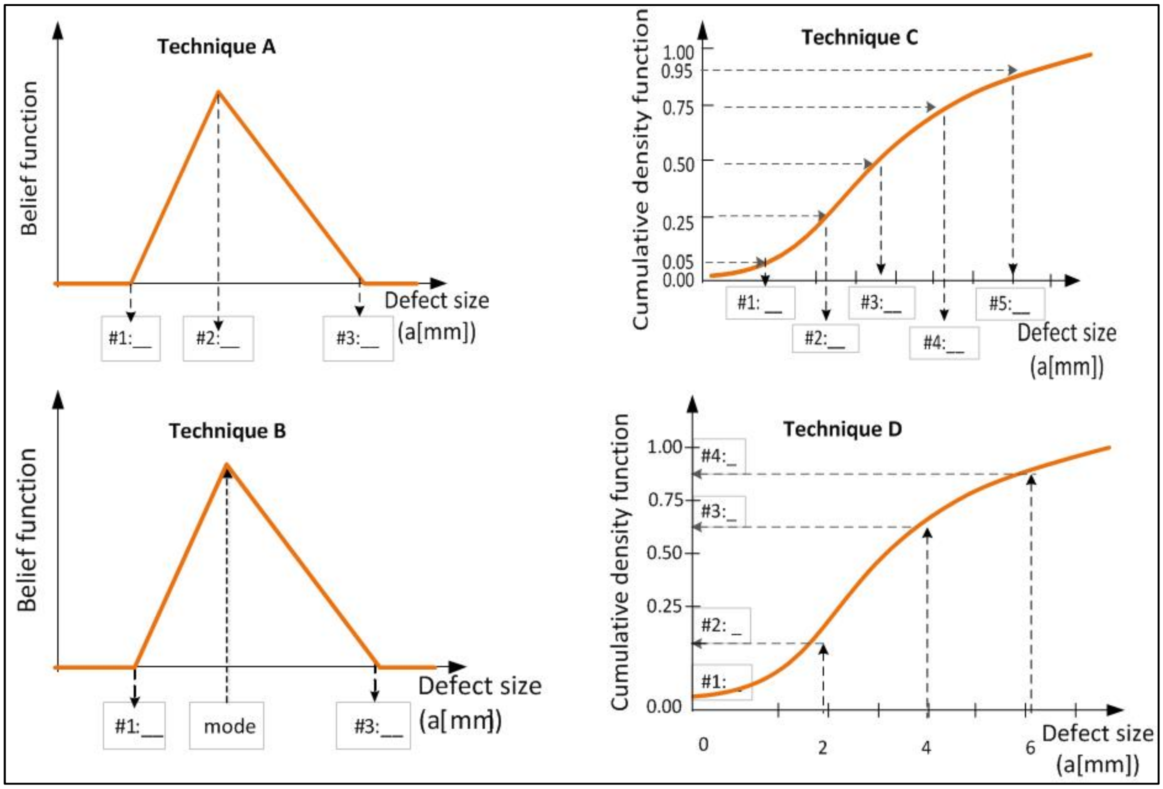


Figure 4.4 Suggested elicitation techniques

4.3.2.1 Technique A

The prediction of defect size $a[mm]$ can present epistemic uncertainty and expressing it as a single value will be inappropriate since it will not include information on the uncertainties related to expert knowledge (Meyer et Booker, 2001). In technique A (Figure 4.4), the expert is invited to formulate his opinion as a bounded interval (min, max) with the most probable value ($mode$). This formulation can be assimilated to a triangular-possibility shape and is very intuitive for experts. This technique can be classified under the direct technique category and its purpose is to assess the effect of minimal information on expert opinion.

4.3.2.2 Technique B

In this technique, we provide a support to the expert and suggested an expected value of the defect size ($a = 2\ mm$) for all operating conditions (Figure 4.4). The expert can express his opinion by providing the minimum (min) and maximum (max) values around the proposed value. This suggested value (anchoring value) is provided to the expert in order to guide him in formulating his opinion, but on other hand this support could introduce some bias to his opinion; a risk which we intend to evaluate through this technique.

4.3.2.3 Technique C

Certain studies suggest combining elicitation techniques like estimating physical quantity jointly with its associated probability (Meyer et Booker, 2001; NUREG-1150, 1990). However, providing simultaneously two sets of data may be hard to manage by the experts. In technique C, we ask experts to assign the likely defect size value to each suggested percentile in the proposed list, such that:

$$P(X < x) = p \tag{4.1}$$

Where X is the quantity for which the expert estimates probabilities, and p represents the provided probability value in the proposed list (e.g. 5%, 25%, 50%, 75% and 95%) as shown in Figure 4.4.

4.3.2.4 Technique D

In technique D (Figure 4.4), we suggest a list of possible defect sizes for which experts are asked to provide suitable probability p for each provided defect size x, such that:

$$P(X < x) = p \tag{4.2}$$

Through this multi-values list, we intend to help experts to focus on the relevance of each defect size and then formulate the suitable probability. The choice of this technique is supported by other studies which demonstrated that for cases with scarce information, people make better indirect judgments than direct estimates (Meyer et Booker, 2001).

4.3.3 Elicitation techniques comparison

In this study, we investigated the differences between elicitation techniques, which was performed based on the collected expert opinions for each elicitation technique (A, B, C &

D). The observed difference could be measured through statistical tests such as the mean *t-Student* test; however these tests do not provide a quantitative measure of the agreement between experts. For this purpose, we propose the use of the Divergence Metric (DM) $\delta(O_i, O_m)$, which allows the measurement of the divergence between two different experts E_i and E_m as follows:

$$\delta(O_i, O_m) = \int_{\mathbf{\Omega}} \left| F_{O_i}(x) - F_{O_m}(x) \right| dx \tag{4.3}$$

The expert E_i provides the opinion $O_i \sim F_i$ according to the distribution F_i and the expert E_m provides the opinion $O_m \sim F_m$ according to the distribution F_m. $\mathbf{\Omega}$ represents the common support domain for both expert opinions.

The elicitation process, discussed in Section 4.3.1, is a multi-effects problem where we intend to assess the influence of elicitation techniques on expert opinions and at the same time to evaluate if the operating conditions change influences expert opinions. Consequently, the provided opinions $O_{i,j}^k$ have 3 dimensions: i represents the expert number ($i = 1, \dots, N$), j is the technique number in $\{A, B, C, D\}$ and k is the operating condition (Case # k, Tableau 4.1). The methodology used is presented in Figure 4.5.

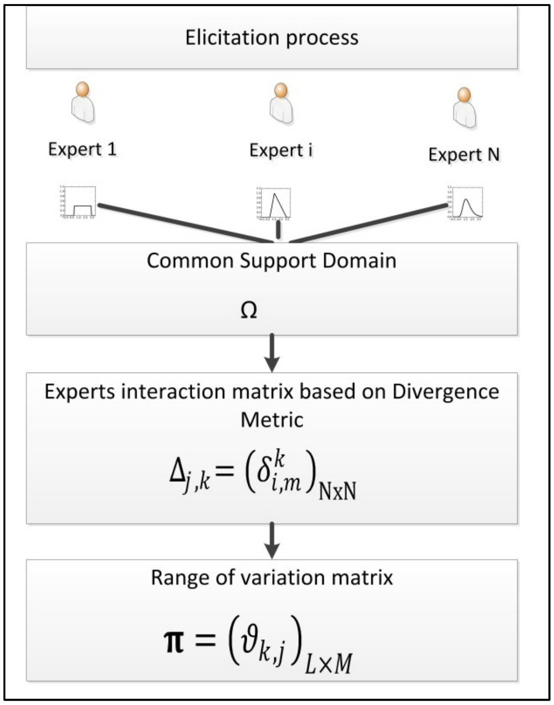


Figure 4.5 Schema of the proposed algorithm for the elicitation technique comparison

In this methodology, each expert E_i from the N available experts provides an opinion $O_{i,j}^{k}$ for the elicitation technique (j) and the operating condition (k). This opinion is supposed to be a random variable (RV) following a specific distribution as suggested by the elicitation technique $O_{i,j}^{k} \sim F_{i,j}^{k}$. The main steps of this methodology are summarized as follows:

- Definition of a common support domain for all provided opinions: $\boldsymbol{\Omega} = [\min O_{i,j}^{k}, \max O_{i,j}^{k}]$ where $i = \{1, \dots, N\}$, $j \in [A, B, C, D]$, $k = \{1, \dots, L\}$ is the operating condition, N is the number of experts available ($N = 5$ in our study), M is the number of techniques ($M = 4$) and L is the number of operating conditions ($L = 4$ as proposed in Tableau 4.1).
- Complete the matrix $\boldsymbol{\Pi}_i = \left(a_{i,j}^{k}\right)_{L\times M}$ with each expert (E_i) opinion ($O_{i,j}^{k}$) converted to a cumulative distribution, as follows: $a_{i,j}^{k} = \int_{\boldsymbol{\Omega}} F_{i,j}^{k}(x)dx$, for the elicitation technique j and the operating condition k.

- For each elicitation technique j (for a total of M elicitation techniques) and each operating condition k (for a total of L operating conditions), build the expert interaction matrix $\Delta_{j,k} = \left(\delta_{i,m}^{k}\right)_{\mathrm{NxN}}$ where $i, m = 1, \ldots, N$ and $\delta_{i,m}^{k} = \left|a_{i,j}^{k} - a_{m,j}^{k}\right|$. $\delta_{i,m}^{k}$ is the DM between experts E_i and E_m for elicitation technique j and the operating condition k.
- Generate the 'variation matrix' $\boldsymbol{\pi} = \left(\vartheta_{k,j}\right)_{L\times M}$, where $\vartheta_{k,j} = Max\left(\Delta_{j,k}\right) - Min(\Delta_{j,k})$. This matrix reports the range of variation of elicitation techniques for all operating conditions which also informs us on elicitation techniques effect on expert opinion.

4.3.4 Expert opinion assessment

The use of experts' opinions in risk analysis often results from two needs: the lack of information on a given subject and the need to improve decision-making. In such situations, several experts can be designated in order to maximize the information available. Consequently, an adequate aggregation rule should be defined in order to combine the resulting expert opinions. In this study, we propose the use of the Cumulative Distribution Averaging (CDA) since it seems suitable even for opinions expressed according to a non-probabilistic distribution. Thus for each element x_0 in the common support domain $\boldsymbol{\Omega}$, the CDA is defined as a combination of the N experts E_i opinions as follows:

$$F_{CDA}(x_0) = \sum_{i=1}^{N} \omega_i F_i(x_0) \tag{4.4}$$

Where F_i represents the distribution used by the expert E_i to model his opinion. The weight ω_i represents the importance of each distribution F_i (the weight of each expert opinion). In our case, all experts have the same weight $\omega_i = N^{-1}$ which means that the proposed CDA represents the arithmetic mean of expert opinions.

For the studied problem, the F_{CDA} will be defined for each elicitation technique j and each operating condition k. So if expert E_i provides an opinion $O_{i,j}^{k} \sim F_{i,j}^{k}$, then

$$F_{CDA,j}^{k}(x_0) = \sum_{i=1}^{N} \omega_i F_{i,j}^{k}(x_0) \tag{4.5}$$

The opinions provided by the experts for a given variable are not systematically identical which could result in opinion dispersion. This behaviour can be captured and quantified using the *leverage* concept. In the literature, the *leverage* concept has been considered in regression analysis as a tool to measure how far the observations are from the barycenter of the fitted data (Cardinali, 2013; Hoaglin et Welsch, 1978). In our case, we define for each expert E_i the *leverage coefficient* l_i as follows:

$$l_i = \frac{1}{N} + \frac{\delta(O_i, CDA)^2}{\sum_{m=1}^{N} \delta(O_m, CDA)^2} \tag{4.6}$$

Where N is the total expert number, O_i is the expert E_i opinion expressed according to the distribution F_i and δ is the DM described in section 4.3.3.

4.4 Results

Results obtained from the elicitation process are shown in Figure 4.6. In this figure, we note that some experts provide very conservative opinions resulting in marked differences between in opinions (e.g. Expert E_5). This is why in certain figures, only a fraction of the expert's opinion is shown.

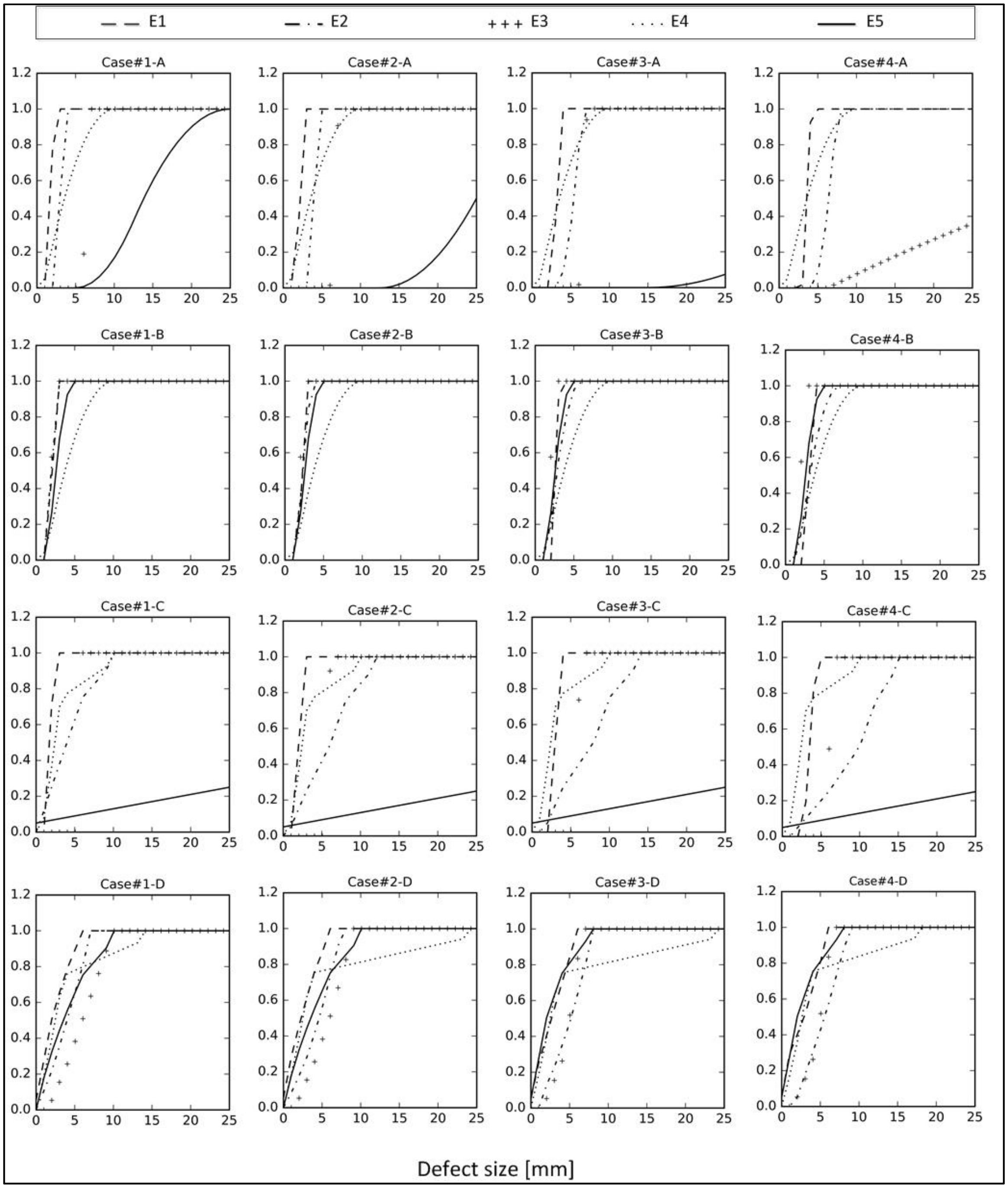


Figure 4.6 Cumulative distribution as function of defect size [mm], representing opinions provided by the five experts (E_i), according to the four elicitation techniques and the four operating conditions

In Figure 4.7, we present the divergence between experts for the elicitation technique A. These results are based on the expression of the Divergence Metric δ presented in section

4.3.3, which gives a quantitative value characterizing the divergence, instead of relying only on the visual aspect of the results shown in Figure 4.6. For example, we note that experts E_1 and E_2 are closer to E_3 (since their DM, $\delta \approx 0$) for all operating conditions except for Case#4. On the other hand, we note that expert E_5 presents the highest divergence among the five experts.

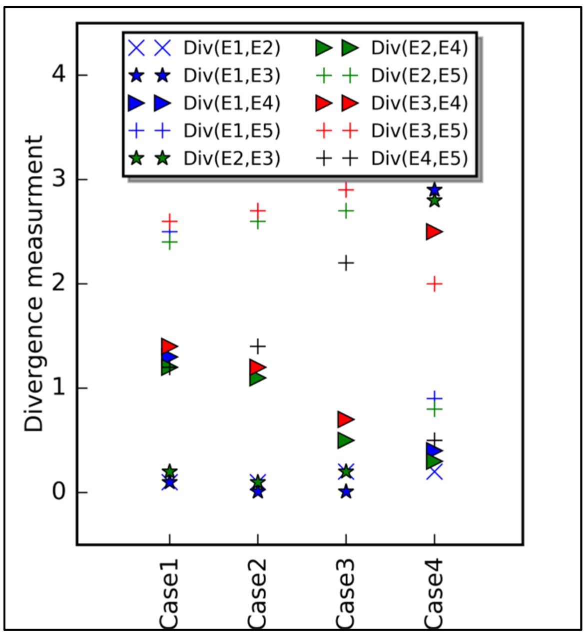


Figure 4.7 Divergence between expert opinions (E_i, E_j) for technique A and for the four operating conditions

In Figure 4.8, the behaviour of the five experts is shown with respect to the group average obtained using the CDA (Eq. 3) for each operating condition (Cases #1 - #4). In this figure, the DM has been used conjointly with the sign of the difference between expert opinion distribution and the CDA.

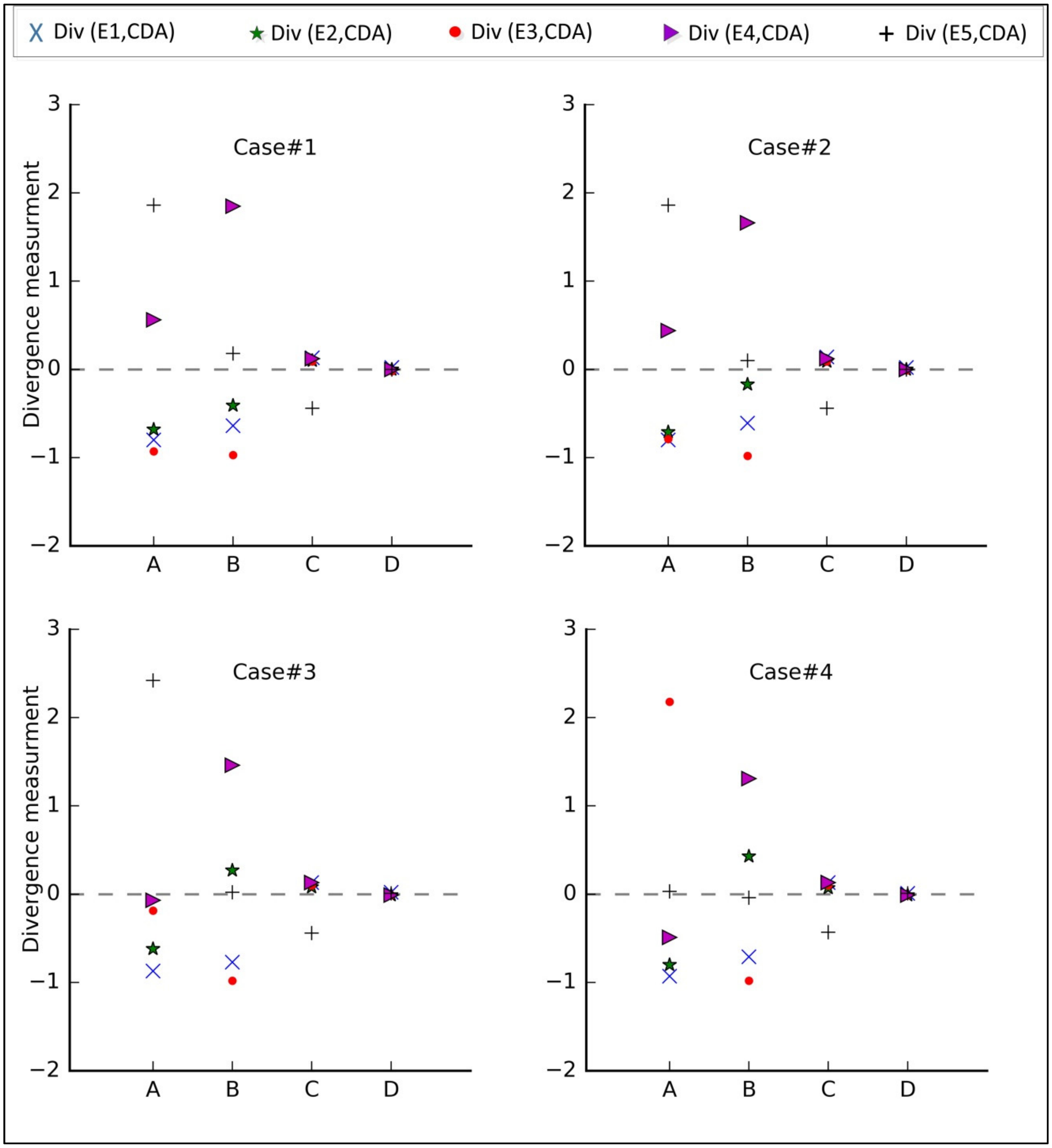


Figure 4.8 Divergence between expert opinions with the CDA of the group, for the four operating conditions and the four elicitation techniques A, B, C and D

We can observe in Figure 8 that for all elicitation techniques and for all operating conditions (except operating conditions Case #4), experts E_1, E_2 and E_3 are closer to each other and to the CDA of the group with a DM $\delta \approx 0$. However, for technique A and technique B, in the operating conditions Case #1-Case #3, experts E_4 and E_5 are always far from the group's average. Furthermore, these results show that experts E_4 and expert E_5 have a significant *leverage* compared to the group average, which is also supported by Figure 4.9. This indicates that these expert opinions will significantly influence the resulting aggregation. In Figure 4.9, we also observe that expert E_4 has a great influence, especially in technique B, and expert E_3 has a noticeable effect in technique A for the operating condition Case #4. Consequently, for such situations, the *leverage coefficient* can be used as a mean to quantify the dispersion between expert opinions.

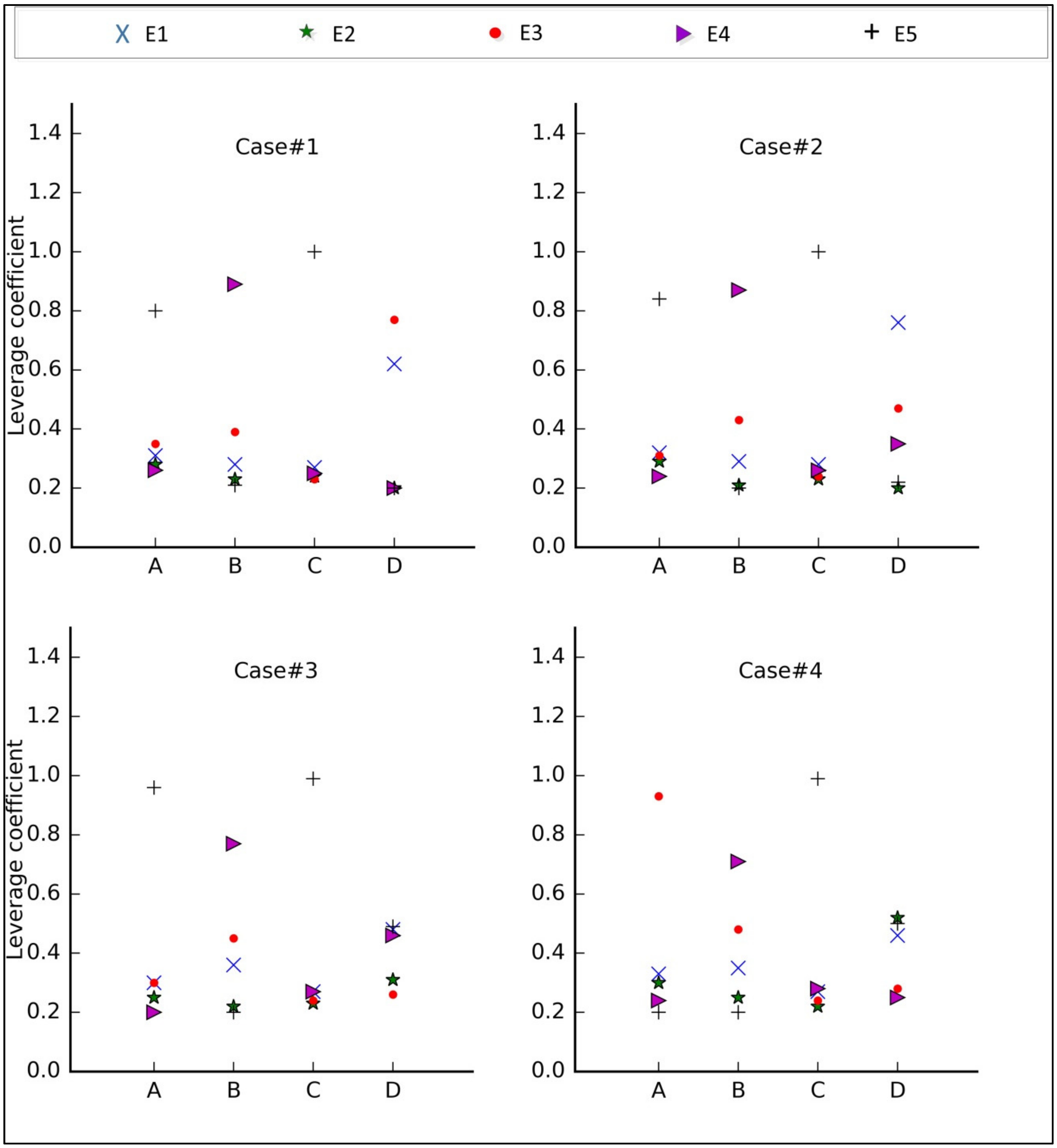


Figure 4.9 Leverage coefficient for the five experts and for the four elicitation techniques and the four operating conditions

4.5 Discussion

From previous results, we observe that the DM δ can be adopted as a tool to measure the agreement and the *leverage coefficient* can be used as a mean to quantify expert opinions dispersion. Thus, based on these two metrics, we can classify experts according to their consistency, which can be translated as the maintaining of the similar 'variation' found between two different operating conditions, on all (or most) elicitation techniques. Consequently this classification might help in defining 'equivalent' experts and in selecting new experts for future elicitation processes. Figure 4.8 and Figure 4.9 also show expert consistency and reproducibility for different elicitation techniques and different operating conditions. In these figures, we can observe the inconsistency between expert E_5 and expert E_3, for particular operating conditions (Figure 4.8, Case #4). This behaviour illustrates the divergence between the two experts' opinions regarding the effect of operating conditions on the defect size prediction.

The elicitation techniques comparisons were performed based on the approach proposed in section 4.3.3. The obtained results are summarized in Tableau 4.2 and Figure 4.10. According to these results, it appears that technique C and technique D are the ones who generate the least variation among experts. For technique D, experts received a list of possible defect values and they were asked to use their knowledge to formulate appropriate probabilities for each proposed 'defect value' in the list. This framework helps experts to focus more on the given values and consequently to provide their best judgements. But, in the absence of sufficient information, experts can assume that the provided values are 'the most likely' for the given operating conditions; a belief that becomes significant if experts do not have enough knowledge to bridge between the provided supports (in the elicitation technique) and the available data. In such conditions, experts will formulate their opinions around the given reference values, leading to comparable expert opinions and thus explaining the low variation observed for the technique D. This behaviour was supported by other studies on the human cognitive mind which indicate that for solving complex problems, people tend to take short cuts and to start with a first impression (or anchor value) to formulate their opinions (Hogarth, 1981; Meyer et Booker, 2001). However, the results obtained with technique D do not mean that this technique is the best one, or that is the technique providing the most likely values of the defect size, but rather it just indicates that this technique generates less variation between experts. So in situations where we have sufficient prior knowledge on the required data, an elicitation framework according to technique D can be suggested in order to limit the variation between experts because this technique allows them to put more focus and effort on the relevancy of each suggested value. Otherwise, and outside of this particular scope, following technique D can distort the 'true' expected values, specifically if the studied system does not have sufficient data in its historical database.

Tableau 4.2 Variation matrix **π** (section 4.3.3) of elicitation techniques A, B, C and D for operating conditions Case #1-Case #4

Technique	Case #1	Case #2	Case #3	Case #4
A	803,0	867,4	1657,2	2743,4
B	426,6	417,7	426,3	411,9
C	30,5	31,1	30,0	29,8
D	3,5	2,9	2,6	2,6

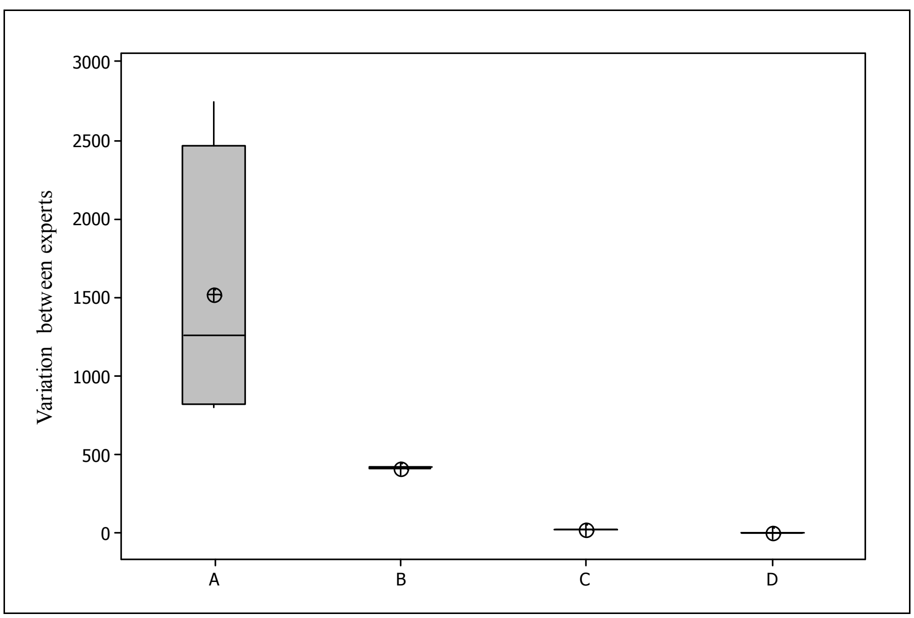


Figure 4.10 Variation in expert opinions (according to the approach described in section 4.3.3) for elicitation techniques A, B, C and D

Results obtained from techniques A and B, do not follow the previous observations. In fact, in technique A, experts seem to have more 'freedom' to express their opinions, which has been reflected by a large variation in expert opinions (Figure 4.10). On other hand, according to the adopted elicitation process, experts performed firstly technique A, followed by technique B which was designed in a similar way to technique A, the only exception is that in technique B we provided the mode value. As the structure of both techniques is similar, some experts attempted to mimic technique A's opinion by simultaneously respecting technique B's requirements. This attitude resulted in comparable opinions in both techniques for some experts, but for others the way the elicitation process was conducted (technique A before technique B) has created for them a certain discomfort, since their opinions for technique A were far from the proposed mode value in technique B.

Consequently, from these results we can state that the elicitation technique can indeed influence expert opinion, an observation which is also shared by other researches (Davis *et al.*, 2006; O'Leary *et al.*, 2009). More, how elicitation process was conducted can also affect elicitation results. Also, through this analysis, we observed that the choice of the appropriate elicitation technique cannot be performed based only on the least variation between experts' criteria, because as illustrated in technique D, in some conditions it can yield to unrealistic results. Our opinion is that the optimal technique should generate less variation between experts and take advantage of expert knowledge by letting him feel comfortable with the use of the proposed technique.

During this study, we noticed that some experts prefer to be conservative instead of providing their best estimation; an ascertainment which is also shared with other researches (Babuscia et Cheung, 2014). This attitude can distort the analysis and the resulting decision, thus experts should be sensitized on the consequences of this behaviour and provide more realistic opinions in order to reach objective decisions. Also, when designing elicitation techniques, the assumption that most experts are comfortable with handling certain mathematical concepts such as probability concepts might be wrong (Babuscia et Cheung, 2014). In fact, in our study, this aspect has represented a challenge to some experts and it had required a quick refresher to minimize misunderstandings that could lead to erroneous information and could increase the expert opinions dispersion. For such situations, Booker *et al.* advocate "extracting the knowledge in as raw (or perhaps pure) and unbiased form as possible, according to the way experts think and problem solve", because the more experts are comfortable with a specific formulation, the more easily they can model their knowledge (Booker et McNamara, 2004). Therefore, it would be appropriate for some cases to substitute probability by a weighting of the proposed events or to guide experts to express their opinion according to a well-defined probabilistic language.

4.6 Conclusion

This study highlighted the difficulty associated to the expert opinions elicitation, in a context with few reference data and operating history information. For such conditions, we proposed and compared four elicitation techniques to support experts with suitable frameworks for their prediction of the defect size, in hydraulic turbine blades. These techniques differ in their structures and in the level of support given to the experts. Findings from this study showed that the choice of the elicitation technique can indeed influence expert opinion; which confirms the observations made by other researchers (Davis et al., 2006; O'Leary et al., 2009). In addition, our results demonstrated that some elicitation techniques can reduce the variation between expert opinions and these techniques are suitable only if prior knowledge on the expected data is available.

In this paper, we demonstrated that the proposed DM δ can be adopted as a tool to measure the agreement between experts and that the leverage coefficient can be used as a mean to quantify expert opinions dispersion. Thus, with these two metrics, experts could be classified according to their consistency in the elicitation process. On other hand, the proposed methodology for the comparison between elicitation techniques could also be adopted conjointly with an appropriate ranking strategy, to reach an objective consensus when such agreement is required.

Finally, during the elicitation process, we noticed that some experts have a 'conservative' attitude which can considerably affect the decision-making process. We also noted that the way the elicitation process has been conducted can in some cases influence expert opinions, therefore we recommend taking into account these factors in order to improve elicitation process results.

4.7 Acknowledgements

The authors would like to give special thanks to the experts for their involvement in the elicitation process and their active participation. Also the authors would like to thank MITACS-Canada and Institut de recherche d’Hydro-Québec (IREQ) for their financial support of the project.

CONCLUSION

Dans cette thèse, nous nous sommes intéressés à l'amélioration de la prédiction de la fiabilité résiduelle en fatigue des turbines hydrauliques et qui est calculée selon le modèle proposé par Gagnon *et al* (Gagnon et *al.*, 2013). L'analyse des paramètres impliqués dans ce modèle a montré que le type de chargement, l'environnement et les modes des réparations ont un impact sur les propriétés micro-structurelles des matériaux, ce qui affecte les propriétés mécaniques des turbines et par conséquent, le taux de propagation des fissures. La connaissance *précise* du comportement de chacun de ces paramètres permettra donc de mieux estimer ses incertitudes (systématiques et épistémiques) et d'améliorer ainsi notre capacité à prédire la fiabilité des turbines. Cependant, l'accès à des données fiables et à jour pour les entrants de ce modèle est souvent difficile dû aux contraintes d'exploitation et d'où le recours aux avis des experts.

Dans cette thèse, nous avons commencé par évaluer certaines théories de modélisation des avis des experts. Ensuite, on a proposé et comparé certaines techniques d'élicitation, en vue d'identifier les plus appropriées au système étudié.

Nous avons élaboré, dans le cadre de cette thèse, des réponses scientifiques aux questions suivantes:

- Comment les théories de probabilité classique et de probabilité imprécise affectent le calcul de la fiabilité?
- Quel est le niveau de reproductibilité de différentes techniques d'élicitation et comment influencent-elles les avis des experts?
- Existe-il une différence entre la combinaison des avis des experts avant leur propagation dans le modèle du système et leur combinaison après leur propagation dans le modèle du système?

Aussi, nous conjecturons que l'ensemble des résultats documentés dans les trois articles produits à cet effet peuvent s'appliquer à d'autres domaines où les avis des experts sont sollicités. L'adoption des résultats avancés dans ces recherches permettra d'améliorer le calcul de fiabilité et de considérer des mesures correctives anticipatoires en conséquence. Ce qui permettra d'optimiser les coûts et modes d'exploitation des parcs hydrauliques. Les stratégies proposées pour le traitement des données d'entrée du modèle de fiabilité peuvent être également adoptées dans d'autres modules du projet PréDDIT d'Hydro-Québec.

Contributions et apports

Dans cette recherche, et dans l'objectif d'une meilleure estimation de la fiabilité, nous avons mis en évidence comment les avis des experts peuvent servir pour la *mise à jour* et pour la prédiction de certaines données. Le CHAPITRE 2 présente nos résultats de l'article #1, sur la comparaison des différentes théories de modélisation des avis des experts.

Les faits saillants de cet article #1 sont comme suit :

- Proposition d'une approche basée sur l'échantillonnage '$\alpha - cut$', permettant l'extension du calcul de fiabilité selon la méthode de l'indice de Hasofer-Lind aux variables issues de probabilité imprécise.
- Suggestion d'une approche qui imite les distributions non-bornées, pour contourner les limitations dues à l'approche FORM.
- Dans cette étude on a conclu que les théories aboutissent à des résultats pratiquement équivalents et que des efforts supplémentaires devront être plus déployés dans la formulation des modèles de propagation et dans l'élicitation des avis des experts. Aussi, nous avons préconisé que les avis soient formulés à l'aide de distributions continues et non bornées.

Le choix du mode d'agrégation (avant propagation ou après propagation) approprié pour un domaine défini, n'est pas clairement identifié dans la littérature. Dans le CHAPITRE 3, nous avons exploré les raisons pouvant expliquer la différence entre les deux modes d'agrégation. Pour le modèle étudié on a noté que cette différence commence à être significative seulement quand on est dans la 'zone sûre' du modèle de fiabilité, proche de son état limite.

Les principaux faits saillants de l'article#2 sont comme suit :

- L'emplacement du point de fonctionnement et les écarts entre les avis d'experts sont les principaux facteurs expliquant la différence entre les deux modes d'agrégation.
- Nous avons proposé une métrique de divergence (δ) basée sur le concept du *Area Metric,* pour quantifier la divergence entre les experts.
- La règle d'agrégation basée sur le concept '*Cumulative Distribution Averaging* CDA' permet d'éliminer les limitations engendrées par les distributions bornées, dans le calcul de fiabilité des modèles basés sur l'approche FORM.

Dans le CHAPITRE 4, nous avons abordé le problème des techniques d'élicitation des avis des experts. Nous avons comparé quatre différentes techniques d'élicitation en vue d'évaluer leurs reproductibilités. Comme corollaire à cette première étape, nous avons tenté, dans l'article #3, de déterminer celles qui s'adaptent mieux à la problématique de fiabilité en fatigue. Les données utilisées dans cette étude proviennent d'un processus d'élicitation empirique qui a été mené exclusivement pour cet effet avec les experts d'Hydro-Québec.

Les principaux résultats de l'article #3, sont comme suit :

- Une étude expérimentale de reproductibilité a été réalisée avec cinq experts et quatre techniques d'élicitation pour quatre conditions d'exploitation différentes. Ce processus d'élicitation a visé la prédiction de la taille des défauts dans les aubes des turbines hydrauliques.
- Suggestion d'une approche pour la comparaison des techniques d'élicitation, basée sur la limitation de variation entre les avis des experts.
- Le coefficient de mesure de divergence δ et le coefficient du levier ont été utilisés comme outils pour mesurer l'accord entre les experts et pour analyser la dispersion des avis d'experts.
- Nous avons conclu que les techniques d'élicitation avec supports permettent de réduire la variation entre les avis des experts, si on dispose d'une connaissance préalable du comportement de la variable à éliciter. Par contre, l'adoption de ce

support dans le cas d'une variable ayant un pauvre historique, risque de biaiser l'avis de l'expert et de le dévier de la valeur réelle attendue.

- La façon avec laquelle le processus d'élicitation a été mené peut affecter la formulation des opinions d'experts et expliquer certains biais.

Finalement, nous avons conclu que les experts doivent être sensibilisés sur l'attitude d'être 'conservateur' parce que ce comportement peut affecter considérablement la décision prise en conséquence.

Originalité de la thèse

L'originalité des travaux effectués a permis de toucher à des sujets qui, selon notre connaissance, n'ont reçu que peu d'intérêt. Dans ce sens on cite: l'effet des théories de modélisation sur les avis des experts, les techniques d'élicitation appropriées pour la fiabilité en fatigue et la différence entre les agrégations avant et après propagation.

Ainsi l'ensemble des publications que nous avons proposées pour résoudre la problématique de recherche de cette thèse se présente comme suit:

- M. Berdaï, A. Tahan, M. Gagnon, "*Imprecise probabilities in fatigue reliability assessment of hydraulic turbines*", ASME J. Risk Uncertainty Part B, September 2016.
- M. Berdaï, A. Tahan, M. Gagnon, "*Comparison between aggregation before propagation and after propagation based on a reliability model*", soumis au Journal Information Sciences.
- M. Berdaï, A. Tahan, M. Gagnon, "*Reproducibility investigation of elicitation techniques based on a fatigue reliability model*", soumis au Journal of Risk Research.
- M. Berdaï, A. Tahan, M. Gagnon "*Maintenance strategy in fatigue based on the Hasofer-Lind reliability index*", 2016: 11th MOSIM Conf. , August 2016

RECOMMANDATIONS

Dans cette thèse nous avons exploré certaines pistes permettant l'amélioration de notre capacité à estimer la fiabilité résiduelle en fatigue des turbines hydroélectrique. Ces travaux ont été plus axés : sur l'incorporation des avis des experts dans un modèle de fiabilité, sur l'effet des techniques d'élicitation sur les avis des experts et sur le mode d'agrégation des avis des experts.

Toutefois, à cause des contraintes temporelles, nous n'avons pas réussi à répondre à toutes les questions générées au cours de cette recherche. Dans cette section, nous présentons certaines de nos réflexions qui peuvent être considérées comme de futurs projets de recherche :

1. Dans l'article #1 (Berdai, Tahan et Gagnon, 2016) nous avons proposé certaines approches permettant d'adapter les avis des experts formulés selon les probabilités imprécises, au modèle de fiabilité étudié. Ainsi, il serait pertinent d'incorporer dans ce modèle de fiabilité, les algorithmes proposés dans l'article#1 afin de permettre au modèle la réception de tous les avis des experts, indépendamment des théories utilisées pour leurs formulations. Cette recommandation s'applique plus précisément aux avis formulés selon des distributions bornées ou basés sur les probabilités imprécises. Par cette étape 'd'uniformisation de traitement' et 'd'ouverture' à plusieurs types d'intrants, on réduit la subjectivité qui peut résulter des traitements nécessaires pour l'adaptation de ces intrants aux exigences du modèle; puisque la formulation de ces traitements peut différer d'un analyste à un autre.
2. Dans le même sens, il serait aussi profitable de connecter le modèle de fiabilité à un modèle d'estimation des contraintes de chargement. Rappelons que pour l'évaluation de la fiabilité résiduelle, on a besoin, entre autres, de la longueur de la fissure et des contraintes de chargement. Dans notre étude, nous nous sommes surtout concentrés sur l'aspect de prédiction de la taille du défaut, par conséquent, l'incorporation des informations sur le chargement, ne sera que bénéfique pour améliorer la précision du calcul par le modèle. À cet égard on cite le récent travail de I. Diagne qui a proposé une

méthodologie pour prédire le niveau des contraintes sur les aubes à partir des données collectées *in situ* et qui présentent une forte corrélation avec les contraintes en question, (Diagne, 2016). Cette méthodologie peut être concrétisée par un modèle qui pourra s'intégrer dans le modèle de fiabilité, pour améliorer les prédictions du modèle de fiabilité.

3. Dans le cadre des travaux de recherche menés dans ce projet, on a proposé une stratégie de maintenance, conditionnée par l'indice de fiabilité et appropriée au problème de fatigue pour les turbines hydrauliques (Berdaï, Tahan et Gagnon, 2016). Comme continuation de ce travail, il serait pertinent de développer un logiciel d'aide à la décision, incorporant un modèle pour la prédiction du moment probable d'atteinte de la défaillance totale des aubes des turbines et incluant aussi les algorithmes discutés dans les points précédents. Avec ce logiciel les gestionnaires des parcs hydrauliques auront la possibilité de planifier le moment optimale pour déclencher l'opération de maintenance, anticipant ainsi la défaillance de la turbine.
4. Le recours aux avis des experts en analyse de risque résulte souvent de deux besoins : le manque d'information sur le sujet étudié et le besoin d'améliorer la prise de décision. Ainsi, les avis des experts sont souvent sollicités pour prédire certaines variables. Généralement, dans ces situations on a un historique limité sur la variable désirée, ou encore une absence complète des valeurs de références. Par conséquent, des techniques d'élicitation appropriées devront être proposées pour supporter les experts dans de tels cas. Dans cet objectif et lors de l'étude de certaines techniques d'élicitation (CHAPITRE 4), nous avons proposé quelques techniques d'élicitation dont certaines contribuent dans la limitation des variations entre les opinions d'experts. Les fonctions multi-attributs proposées par Beaudouin pour l'amélioration de prise de décision peuvent aussi être adoptées pour réaliser cet objectif (Beaudouin, 2015). Cependant ces techniques seront surtout appropriées quand la gamme des valeurs proposées est proche des valeurs réelles. Autrement, l'adoption de ces techniques pourrait biaiser et même fausser les avis des experts, par rapport aux vraies valeurs attendues. Il s'ensuit que le choix de la technique d'élicitation appropriée ne peut être effectué seulement sur la base du critère de la moindre variation entre experts et d'autres critères d'évaluation devront être considérés.

Ainsi, pour la prédiction dans un contexte fortement incertain il serait recommandé d'élaborer une technique d'élicitation permettant une fusion entre les concepts des techniques A et D proposées au CHAPITRE 4.

5. Dans le même objectif, une réplication de la collecte des avis des experts permettrait d'estimer l'erreur de répétitivité et de bâtir une riche base de données sur les valeurs à éliciter. Cette base, pourrait être exploitée pour développer un 'pseudo-modèle' décrivant le comportement de la variable à éliciter et du même coup servir comme moyen de validation des avis des experts et un outil pour la formation des experts.
6. La disponibilité de plusieurs avis d'experts sur différentes périodes (découlant de différent processus d'élicitations), nous amène à se poser la question sur la façon de la mise à jour des avis dont on dispose à priori: est ce que les nouveaux jugements devront '*remplacer*' les jugements précédents? Ou bien les nouveaux jugements doivent '*tenir compte*' des précédents? Dans la revue de littérature on constate que l'inférence Bayésienne reste la stratégie de mise à jour la plus référée dans les évaluations fiabilistes (Friedman, Formichi et Landi, 2017). L'inférence bayésienne révise la probabilité des propositions au fur et à mesure que des observations/ avis soient disponibles. Ainsi l'effet de la distribution a priori s'estompe au fur et à mesure que les observations sont prises en compte. Ce choix implique que les avis à postériori seront une '*révision avec mémoire'* des avis précédents; un choix qui doit être supporté par des arguments solides, surtout dans l'évaluation de fiabilité dans un contexte incertain.

   Par conséquent une étude plus approfondie de ce sujet est nécessaire, pour évaluer d'autres stratégies de mise à jour, telles que les stratégies basées sur les filtres d'information (Hashlamon et Erbatur, 2016) et de choisir une stratégie appropriée pour le modèle de fiabilité en cours d'étude, tenant compte du format de l'avis de l'expert qui peut être sous un format numérique (valeur simple, distribution) ou selon un format linguistique.
7. Les avis des experts peuvent être des valeurs numériques comme ils peuvent être sous une forme linguistique; une forme qui est supportée et recommandée par plusieurs recherches (Booker et McNamara, 2004). Cependant ce type de formulation présente une limitation quand il s'agit de l'évaluation de la fiabilité, où il est essentiel de transformer

les avis ‘linguistiques’ en valeurs numériques; une transformation qui ajoute une subjectivité à l’information requise (Zhang, Mahadevan et Deng, 2017). Par conséquent le développement de transformations ‘standards’ permettra de réduire ces subjectivités. Dans ce sens il convient de citer le travail de (Ramos-Soto, Bugarín et Barro, 2016) qui proposent deux nouvelles branches : ‘*the natural language generation*’ (NLG) et ‘*the linguistic descriptions of data*’ (LDD). Le premier domaine traite du problème général de la conversion des données en textes compréhensibles, tandis que le second se concentre sur l'abstraction des données dans des concepts linguistiques structurés à l'aide d'ensembles flous.

8. Dans l’article #2 nous avons constaté que la règle d’agrégation, basée sur ‘*Cumulative Distribution Averaging* CDA’ élimine les limitations que peuvent engendrer les distributions bornées dans le calcul de fiabilité des modèles basés sur l’approche FORM. Cependant l’effet des autres règles d’agrégation sur le modèle de fiabilité étudié, est méconnu. Nous recommandons une étude comparative des certaines règles d’agrégation pour en choisir la règle la plus appropriée au modèle de fiabilité proposé dans la référence (Gagnon et al., 2013).
9. Finalement, nous conjecturons que les résultats obtenus par nos recherches présentent un potentiel pour être applicables à d’autres domaines de fiabilité, d’analyse de risque et de prise de décision. De ce fait, il serait intéressant de les confirmer en les appliquant sur d’autres modèles de fiabilité.

BIBLIOGRAPHIE

Allan, R. N., et J. Roman. 1989. « Reliability assessment of generation systems containing multiple hydro plant using simulation techniques ». *Power Systems, IEEE Transactions on,* vol. 4, nº 3, p. 1074-1080.

Aoudjit, Hakim. 2010. « Planification de la maintenance d'un parc de turbines-alternateurs par programmation mathématique ». École Polytechnique de Montréal.

ASME, The American Society of Mechanical Engineers. 2012. « Concepts of verification and validation ».

Aven, T. 2012. « Foundations of risk analysis ». < http://site.ebrary.com/id/10529296 >.

Aven, Terje, et Seth Guikema. 2011. « Whose uncertainty assessments (probability distributions) does a risk assessment report: the analysts' or the experts'? ». *Reliability Engineering & System Safety,* vol. 96, nº 10, p. 1257-1262.

Aven, Terje, Baraldi Piero, Roger Flage et Enrico Zio. 2014. « Uncertainty in risk assessment: the representation and treatment of uncertainties by probabilistic and non-probabilistic methods ».

Aven, Terje, Enrico Zio, Piero Baraldi et Roger Flage. 2013. *Uncertainty in risk assessment: the representation and treatment of uncertainties by probabilistic and non-probabilistic methods*. John Wiley & Sons.

Ayyub, Bilal M. 2001. « Elicitation of expert opinions for uncertainty and risks ». *CRC press LLC*.

Babuscia, Alessandra, et Kar-Ming Cheung. 2014. « An approach to perform expert elicitation for engineering design risk analysis: methodology and experimental results ». *Journal of the Royal Statistical Society: Series A (Statistics in Society),* vol. 177, nº 2, p. 475-497.

Baik, Biehn, Kentaro Yamada et Toshiyuki Ishikawa. 2011. « Fatigue crack propagation analysis for welded joint subjected to bending ». *International Journal of Fatigue,* vol. 33, nº 5, p. 746-758.

Baraldi, Piero, Irina Crenguta Popescu et Enrico Zio. 2010. « Methods of uncertainty analysis in prognostics ». *International Journal of Performability Engineering,* vol. 6, nº 4, p. 303-330.

Baudrit, Cédric, et Didier Dubois. 2005. « Comparing Methods for Joint Objective and Subjective Uncertainty Propagation with an example in a risk assessment ». In *ISIPTA*. Vol. 5.

Beaudouin, F. 2015. « Implementing a Multiple Criteria Model to Debate About Nuclear Power Plants Safety Choices ». *Group Decision and Negotiation,* vol. 24, nº 6, p. 1035-1063.

Bech-Larsen, Tino, et Niels Asger Nielsen. 1999. « A comparison of five elicitation techniques for elicitation of attributes of low involvement products ». *Journal of Economic Psychology,* vol. 20, nº 3, p. 315-341.

Beck, André Teófilo. 2016. « Strategies for finding the design point under bounded random variables ». *Structural Safety,* vol. 58, p. 79-93.

Beer, Michael, Scott Ferson et Vladik Kreinovich. 2013. « Imprecise probabilities in engineering analyses ». *Mechanical Systems and Signal Processing,* vol. 37, nº 1–2, p. 4-29.

Ben-Arieh, David, et Zhifeng Chen. 2006. « Linguistic group decision-making: opinion aggregation and measures of consensus ». *Fuzzy Optimization and Decision Making,* vol. 5, nº 4, p. 371-386.

Berdaï, Mounia, Antoine Tahan et M Gagnon. 2016. « Maintenance strategy in fatigue based on the Hasofer-Lind reliability index ». *Conférence MOSIM 2016,* vol. 11ème conférence francophone de modélisation, optimisation et simulation.

Berdai, Mounia, Antoine Tahan et Martin Gagnon. 2016. « Imprecise probabilities in fatigue reliability assessment of hydraulic turbines ». *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part B: Mechanical Engineering.*

Bergmann-Paulsen, Jonas. 2012. « FSI-analysis of a Francis turbine ». Norwegian University of Science and Technology.

Booker, Jane M, et Laura A McNamara. 2004. « Solving black box computation problems using expert knowledge theory and methods ». *Reliability Engineering & System Safety,* vol. 85, nº 1, p. 331-340.

Bourdon, Paul, Mohamed Farhat, Youssef Mossoba et Pierre Lavigne. 1999. « Hydro turbine profitability and cavitation erosion ». In *Waterpower'99@ sHydro's Future: Technology, Markets, and Policy*. p. 1-10. ASCE.

Breivik, Einar, et Magne Supphellen. 2003. « Elicitation of product attributes in an evaluation context: A comparison of three elicitation techniques ». *Journal of Economic Psychology,* vol. 24, nº 1, p. 77-98.

Brekke, Hermod. 2013. « Design, Performance and Maintenance of Francis Turbines ». *Global Journal of Researches In Engineering,* vol. 13, nº 5.

Brickstad, Bjorn, et BL Josefson. 1998. « A parametric study of residual stresses in multi-pass butt-welded stainless steel pipes ». *International Journal of Pressure Vessels and Piping,* vol. 75, nº 1, p. 11-25.

BSI, BS. 2005. « 7910: Guide on methods for assessing the acceptability of flaws in metallic structures ». *British Standards Institute*.

Canfield, H.-R. Bae · R. V. Grandhi · R. A. 2006. « Sensitivity analysis of structural response uncertainty propagation using evidence theory ». *structural and multidisciplinary optimization,* vol. 31, nº 4, p. 270-279.

Cardinali, Carla. 2013. « Observation influence diagnostic of a data assimilation system ». In *Data Assimilation for Atmospheric, Oceanic and Hydrologic Applications (Vol. II)*. p. 89-110. Springer.

Castillo, E., A. Fernández-Canteli, C. Castillo et C. M. Mozos. 2010. « A new probabilistic model for crack propagation under fatigue loads and its connection with Wöhler fields ». *International Journal of Fatigue,* vol. 32, nº 4, p. 744-753.

Castillo, Enrique, Alan J O'Connor, María Nogal et Aida Calviño. 2014. « On the physical and probabilistic consistency of some engineering random models ». *Structural Safety,* vol. 51, p. 1-12.

Chen, Li, et Pearl Pu. 2004. *Survey of preference elicitation methods*.

Chen, Ting-Yu, Hsiao-Pin Wang et Yen-Yu Lu. 2011. « A multicriteria group decision-making approach based on interval-valued intuitionistic fuzzy sets: A comparative perspective ». *Expert Systems with Applications,* vol. 38, nº 6, p. 7647-7658.

Clemen, Robert T., et Robert L. Winkler. 1999. « Combining Probability Distributions From Experts in Risk Analysis ». *Risk Analysis,* vol. 19, nº 2.

Cobb, Barry R, et Prakash P Shenoy. 2006. « On the plausibility transformation method for translating belief function models to probability models ». *International Journal of Approximate Reasoning,* vol. 41, nº 3, p. 314-330.

Cojazzi, G, G Guida, L Pinola, R Sardella et P Baroni. 1987. « KEEJAM: A knowledge engineering methodology for expert judgment acquisition and modeling in probbabilistic safety assessment ». *Advances in Safety and Reliability,* vol. 1, p. 199-206.

Cooke, Nancy J. 1994. « Varieties of knowledge elicitation techniques ». *International Journal of Human-Computer Studies,* vol. 41, nº 6, p. 801-849.

Cooke, Roger. 2004. « The anatomy of the squizzel: the role of operational definitions in representing uncertainty ». *Reliability Engineering & System Safety,* vol. 85, nº 1, p. 313-319.

Cooke, Roger M, et LJH Goossens. 1999. « Procedures guide for structured expert judgment ». *Brussels-Luxembourg: Commission of the European Communities*.

Cooke, Roger M, et Louis HJ Goossens. 2004. « Expert judgement elicitation for risk assessments of critical infrastructures ». *Journal of Risk Research,* vol. 7, nº 6, p. 643-656.

Cooke, Roger M., et Louis L. H. J. Goossens. 2008. « TU Delft expert judgment data base ». *Reliability Engineering & System Safety,* vol. 93, nº 5, p. 657-674.

Dabayeh, AA, et TH Topper. 1995. « Changes in crack-opening stress after underloads and overloads in 2024-T351 aluminium alloy ». *International Journal of Fatigue,* vol. 17, nº 4, p. 261-269.

Davis, Alan, Oscar Dieste, Ann Hickey, Natalia Juristo et Ana M Moreno. 2006. « Effectiveness of requirements elicitation techniques: Empirical results derived from a systematic review ». In *Requirements Engineering, 14th IEEE International Conference*. p. 179-188. IEEE.

Deng, Dean, et Hidekazu Murakawa. 2006. « Numerical simulation of temperature field and residual stress in multi-pass welds in stainless steel pipe and comparison with experimental measurements ». *Computational materials science,* vol. 37, nº 3, p. 269-277.

Destercke, Sébastien, Didier Dubois et Eric Chojnacki. 2006. « Aggregation of expert opinions and uncertainty theories ». In *Rencontres Francophones sur la Logique Floue et ses Applications*. p. 295-302. Cepadues.

Diagne, Ismahel Aboubacar. 2016. « Optimisation des contraintes mécaniques au démarrage des trubines hydroélectriques à l'aide de mesures indirectes ». École de technologie supérieure.

Dieste, Oscar, et Natalia Juristo. 2011. « Systematic review and aggregation of empirical studies on elicitation techniques ». *Software Engineering, IEEE Transactions on,* vol. 37, nº 2, p. 283-304.

Dubois, Didier, Laurent Foulloy, Gilles Mauris et Henri Prade. 2004. « Probability-possibility transformations, triangular fuzzy sets, and probabilistic inequalities ». *Reliable computing,* vol. 10, nº 4, p. 273-297.

Dubois, Didier, et Henri Prade. 1998. « Possibility theory: qualitative and quantitative aspects ». In *Quantified representation of uncertainty and imprecision*. p. 169-226. Springer.

Dubois, Didier, Henri Prade et Philippe Smets. 2008. « A definition of subjective possibility ». *International Journal of Approximate Reasoning,* vol. 48, nº 2, p. 352-364.

Fallet, G., C. Duval, C. Simon, P. Weber et B. Iung. 2011. « Expert judgments collecting and modeling: Application to the Integrated Risks Analysis (IRA) methodology ». In *Dependable Control of Discrete Systems (DCDS), 2011 3rd International Workshop on*. (15-17 June 2011), p. 72-77.

Fatemi, Ali, et Darrell F Socie. 1988. « A Critical Plane Approach to Multiaxial Fatigue Damage Including out-of-Phase Loading ». *Fatigue & Fracture of Engineering Materials & Structures,* vol. 11, nº 3, p. 149-165.

Ferson, Scott, Cliff A Joslyn, Jon C Helton, William L Oberkampf et Kari Sentz. 2004. « Summary from the epistemic uncertainty workshop: consensus amid diversity ». *Reliability Engineering & System Safety,* vol. 85, nº 1, p. 355-369.

Ferson, Scott, William L. Oberkampf et Lev Ginzburg. 2008. « Model validation and predictive capability for the thermal challenge problem ». *Computer Methods in Applied Mechanics and Engineering,* vol. 197, nº 29–32, p. 2408-2430.

Flage, Roger, Piero Baraldi, Enrico Zio et Terje Aven. 2013. « Probability and Possibility-Based Representations of Uncertainty in Fault Tree Analysis ». *Risk Analysis,* vol. 33, nº 1, p. 121-133.

Friedman, Noemi, Paolo Formichi et Filippo Landi. 2017. « Seismic Reliability Assessment of a Concrete Water Tank Based on the Bayesian Updating of the Finite Element Model ».

Frunzăverde, D, S Muntean, G Mărginean, V Câmpian, L Marşavina, R Terzi et V Şerban. 2010. « failure analysis of a francis turbine runner ». *IP science*.

Gagnon, Martin. 2013. *Contribution à l'évaluation de la fiabilité en fatigue des turbines hydroélectriques*.

Gagnon, Martin, Antoine Tahan, Philippe Bocher et Denis Thibault. 2013. « A probabilistic model for the onset of High Cycle Fatigue (HCF) crack propagation: Application to hydroelectric turbine runner ». *International Journal of Fatigue,* vol. 47, nº 0, p. 300-307.

Grooteman, Frank. 2008. « A stochastic approach to determine lifetimes and inspection schemes for aircraft components ». *International Journal of Fatigue,* vol. 30, nº 1, p. 138-149.

Ha-Duong, Minh. 2008. « Hierarchical fusion of expert opinions in the Transferable Belief Model, application to climate sensitivity ». *International Journal of Approximate Reasoning,* vol. 49, nº 3, p. 555-574.

Hahn, Gerald J, et Samuel S Shapiro. 1968. « Statistical models in engineering ». In *Statistical models in engineering*. John Wiley & Sons.

Hashlamon, Iyad, et Kemalettin Erbatur. 2016. « An improved real-time adaptive Kalman filter with recursive noise covariance updating rules ». *Turkish journal of electrical engineering & computer sciences,* vol. 24, nº 2, p. 524-540.

Helton, Jon C., et Jay D. Johnson. 2011. « Quantification of margins and uncertainties: Alternative representations of epistemic uncertainty ». *Reliability Engineering & System Safety,* vol. 96, nº 9, p. 1034-1052.

Herrera-Viedma, Enrique, Francisco Herrera et Francisco Chiclana. 2002. « A consensus model for multiperson decision making with different preference structures ». *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on,* vol. 32, nº 3, p. 394-402.

Hoaglin, David C, et Roy E Welsch. 1978. « The hat matrix in regression and ANOVA ». *The American Statistician,* vol. 32, nº 1, p. 17-22.

Hogarth, Robin M. 1981. « Beyond discrete biases: Functional and dysfunctional aspects of judgmental heuristics ». *Psychological Bulletin,* vol. 90, nº 2, p. 197.

Hora, Stephen C. 1992. « Acquisition of expert judgment: Examples from risk assessment ». *Journal of Energy Engineering,* vol. 118, nº 2, p. 136-148.

Hsu, Hsi-Mei, et Chen-Tung Chen. 1996. « Aggregation of fuzzy opinions under group decision making ». *Fuzzy sets and systems,* vol. 79, nº 3, p. 279-285.

Hudlicka, Eva. 1996. « Requirements elicitation with indirect knowledge elicitation techniques: comparison of three methods ». In *Requirements Engineering, 1996., Proceedings of the Second International Conference on*. p. 4-11. IEEE.

Huth, Hans-Jörg. 2005. « fatigue design of hydraulic turbine runners ». p. 178.

Johannesson, P. 2006. « Extrapolation of load histories and spectra ». *Fatigue & Fracture of Engineering Materials & Structures,* vol. 29, nº 3, p. 209-217.

Kaplan, Stan. 1992. « 'Expert information'versus 'expert opinions'. Another approach to the problem of eliciting/combining/using expert knowledge in PRA ». *Reliability Engineering & System Safety,* vol. 35, nº 1, p. 61-72.

Kitagawa, Hideo, Ryoji Yuuki et Toshiaki Ohira. 1975. « Crack-morphological aspects in fracture mechanics ». *Engineering Fracture Mechanics,* vol. 7, nº 3, p. 515-529.

Klir, GeorgeJ, et RichardM Smith. 2001. « On Measuring Uncertainty and Uncertainty-Based Information: Recent Developments ». *Annals of Mathematics and Artificial Intelligence,* vol. 32, nº 1-4, p. 5-33.

Kuhnert, Petra M, Tara G Martin et Shane P Griffiths. 2010. « A guide to eliciting and using expert knowledge in Bayesian ecological models ». *Ecology letters,* vol. 13, nº 7, p. 900-914.

Lassen, Tom. 2013. « Risk based Fatigue Inspection Planning–State of the Art ». *Procedia Engineering,* vol. 66, p. 489-499.

Lasserre, Virginie. 1999. « Modélisation floue des incertitudes de mesures de capteurs ». Chambéry.

Limbourg, Philipp, et Etienne De Rocquigny. 2010. « Uncertainty analysis using evidence theory–confronting level-1 and level-2 approaches with data availability and computational constraints ». *Reliability Engineering & System Safety,* vol. 95, nº 5, p. 550-564.

Liu, Yu, Wei Chen, Paul Arendt et Hong-Zhong Huang. 2011. « Toward a better understanding of model validation metrics ». *Journal of Mechanical Design,* vol. 133, nº 7, p. 071005.

Lyu, Chang Hoon, Min Seok Choi, Zhong Yuan Li et Hee Yong Youn. 2010. « Reasoning with imprecise context using improved dempster-shafer theory ». In *Web Intelligence and Intelligent Agent Technology (WI-IAT), 2010 IEEE/WIC/ACM International Conference on*. Vol. 2, p. 475-478. IEEE.

Mannhart, A., A. Bilgic et B. Bertsche. 2007. « Modeling Expert Judgment for Reliability Prediction - Comparison of Methods ». In *Reliability and Maintainability Symposium, 2007. RAMS '07. Annual*. (22-25 Jan. 2007), p. 1-6.

Masson, Marie-Hélène. 2005. « Apports de la théorie des possibilités et des fonctions de croyance à l'analyse de données imprécises ». *Mémoire de direction de recherche*, p. 126.

Meyer, Mary A, et Jane M Booker. 2001. *Eliciting and analyzing expert judgment: a practical guide*, 7. SIAM.

Miller, KJ. 1987. « The behaviour of short fatigue cracks and their initiation Part II-A General summary ». *Fatigue & Fracture of Engineering Materials & Structures,* vol. 10, nº 2, p. 93-113.

Moan, T. 2005. « Reliability-based management of inspection, maintenance and repair of offshore structures ». *Structure and Infrastructure Engineering,* vol. 1, nº 1, p. 33-62.

Möller, Bernd, Wolfgang Graf, Michael Beer et J Sickert. 2001. « Fuzzy probabilistic method and its application for the safety assessment of structures ». In *Proceedings of the European Conference on Computational Mechanics, Cracow, Poland*.

Montibeller, Gilberto, et Detlof Winterfeldt. 2015. « Cognitive and Motivational Biases in Decision and Risk Analysis ». *Risk Analysis*.

NUREG-1150. 1990. « Severe accident risks: an assessment for five US Nuclear Power Plants NUREG1150 ». *US Nuclear Regulatory commission*.

O'Hagan, Michael. 1988. « Aggregating template or rule antecedents in real-time expert systems with fuzzy set logic ». In *Signals, Systems and Computers, 1988. Twenty-Second Asilomar Conference on*. Vol. 2, p. 681-689. IEEE.

O'Leary, Rebecca A, Samantha Low Choy, Justine V Murray, Mary Kynn, Robert Denham, Tara G Martin et Kerrie Mengersen. 2009. « Comparison of three expert elicitation methods for logistic regression on predicting the presence of the threatened brush-tailed rock-wallaby Petrogale penicillata ». *Environmetrics,* vol. 20, nº 4, p. 379.

O'Hagan, Anthony. 2012. « Probabilistic uncertainty specification: Overview, elaboration techniques and their application to a mechanistic model of carbon flux ». *Environmental Modelling & Software,* vol. 36, p. 35-48.

Oberkampf, William L., Jon C. Helton, Cliff A. Joslyn, Steven F. Wojtkiewicz et Scott Ferson. 2004. « Challenge problems: uncertainty in system response given uncertain parameters ». *Reliability Engineering & System Safety,* vol. 85, nº 1–3, p. 11-19.

Onoufriou, Toula. 1999. « Reliability based inspection planning of offshore structures ». *Marine structures,* vol. 12, nº 7, p. 521-539.

Oussalah, Mourad. 2000. « On the probability/possibility transformations: a comparative analysis ». *International Journal Of General System,* vol. 29, nº 5, p. 671-718.

Pompetzki, M. A., T. H. Topper et D. L. DuQuesnay. 1990. « The effect of compressive underloads and tensile overloads on fatigue damage accumulation in SAE 1045 steel ». *International Journal of Fatigue,* vol. 12, nº 3, p. 207-213.

Ramos-Soto, Alejandro, Alberto Bugarín et Senén Barro. 2016. « On the role of linguistic descriptions of data in the building of natural language generation systems ». *Fuzzy Sets and Systems,* vol. 285, p. 31-51.

Richard W. Hertzberg, Richard P. Vinci, Jason L. Hertzberg. 2013. « Deformation and fracture mechanics of engineering materials ». *John Wiley and sons*.

Rizzo, Cesare Mario. 2007. « Application of Reliability Analysis to the Fatigue of Typical Welded Joints of Ships ». *Ship Technology Research,* vol. 54.

Rocquigny, Etienne de. 2012. « Modelling under risk and uncertainty an introduction to statistical, phenomenological and computational methods ». < http://site.ebrary.com/id/10558805 >.

Salehghaffari, Shahab, et Masoud Rais-Rohani. 2013. « Material model uncertainty quantification using evidence theory ». *Mechanical Engineering Science*.

Sanford, Robert Joseph. 2003. *Principles of fracture mechanics*. Prentice Hall New Delhi.

Smets, Philippe. 2007. « Analyzing the combination of conflicting belief functions ». *Information Fusion,* vol. 8, nº 4, p. 387-412.

Suresh, S. 1983. « Crack deflection: implications for the growth of long and short fatigue cracks ». *Metallurgical Transactions A,* vol. 14, nº 11, p. 2375-2385.

Szczota, Mickaël. 2012. « Modélisation de l'historique d'opération de groupes turbine-alternateur ». École de technologie supérieure.

Taylor, D. 1988. « Fatigue thresholds: their applicability to engineering situations ». *International Journal of Fatigue,* vol. 10, nº 2, p. 67-79.

Taylor, David. 1989. *Fatigue thresholds*, 130. Butterworths London.

Thapa, Biraj Singh, Mette Eltvik, Kristine Gjøsæter, Ole G Dahlhaug et Bhola Thapa. 2012. « Design Optimization of Francis Runners for Sediment Handling ». In *Fourth International Conference on Water Resources and Renewable Energy Development in Asia, Thailand*.

Therriault, Daniel, et Marie Bernard. 2013. « Caractérisation des matériaux en fatigue ». *Notes de cours*.

Thibault, D, M Gagnon et S Godin. 2014. « Bridging the gap between metallurgy and fatigue reliability of hydraulic turbine runners ». In *IOP Conference Series: Earth and Environmental Science*. Vol. 22, p. 012019. IOP Publishing.

Thibault, Denis, Philippe Bocher et Marc Thomas. 2009. « Residual stress and microstructure in welds of 13% Cr–4% Ni martensitic stainless steel ». *Journal of Materials Processing Technology,* vol. 209, nº 4, p. 2195-2202.

Thibault, Denis, Martin Gagnon et Stéphane Godin. 2015. « The effect of materials properties on the reliability of hydraulic turbine runners ». *International Journal of Fluid Machinery and Systems,* vol. 8, nº 4, p. 254-263.

Truong, Phuong N, et Gerard Heuvelink. 2013. « Uncertainty quantification of soil property maps with statistical expert elicitation ». *Geoderma,* vol. 202, p. 142-152.

Veritas, Det Norske. 1992. « structural reliability analysis of marine structures ».

Verma, Ajit Kumar, Srividya Ajit et Durga Rao Karanki. 2010. *Reliability and safety engineering*. Springer.

Wall, M, SF Burch et J Lilley. 2009. « Review of models and simulators for NDT reliability (POD) ». *Insight-Non-Destructive Testing and Condition Monitoring,* vol. 51, nº 11, p. 612-619.

Wang, Ying-Ming, et Zhi-Ping Fan. 2007. « Fuzzy preference relations: Aggregation and weight determination ». *Computers & Industrial Engineering,* vol. 53, nº 1, p. 163-172.

Xiao, Ruofu, Zhengwei Wang et Yongyao Luo. 2008. « Dynamic stresses in a Francis turbine runner based on fluid-structure interaction analysis ». *Tsinghua Science & Technology,* vol. 13, nº 5, p. 587-592.

Xiong, J. J., et R. A. Shenoi. 2008. « A load history generation approach for full-scale accelerated fatigue tests ». *Engineering Fracture Mechanics,* vol. 75, nº 10, p. 3226-3243.

Xu, Zeshui, et Qing-Li Da. 2003. « An overview of operators for aggregating information ». *International Journal of Intelligent Systems,* vol. 18, nº 9, p. 953-969.

Xu, ZS. 2004. « Goal programming models for obtaining the priority vector of incomplete fuzzy preference relation ». *International journal of approximate reasoning,* vol. 36, nº 3, p. 261-270.

Yager, Ronald R. 1988. « On ordered weighted averaging aggregation operators in multicriteria decisionmaking ». *Systems, Man and Cybernetics, IEEE Transactions on,* vol. 18, nº 1, p. 183-190.

Yager, Ronald R, et Dimitar P Filev. 1994. « Parameterized AND-UKE and OR-LIKE OWA operators ». *International Journal of General System,* vol. 22, nº 3, p. 297-316.

Yager, Ronald R, et Liping Liu. 2008. *Classic works of the Dempster-Shafer theory of belief functions*, 219. Springer Science & Business Media.

Yuhua, Dong, et Yu Datao. 2005. « Estimation of failure probability of oil and gas transmission pipelines by fuzzy fault tree analysis ». *Journal of Loss Prevention in the Process Industries,* vol. 18, nº 2, p. 83-88.

Zadeh, L. A. 1965. « Fuzzy sets ». *Information and Control,* vol. 8, nº 3, p. 338-353.

Zadeh, L.A. 1978. « Fuzzy sets as a basis for a theory of possibility ». *Fuzzy sets and systems,* vol. 1, p. 3-28.

Zhang, Xiaoge, Sankaran Mahadevan et Xinyang Deng. 2017. « Reliability analysis with linguistic data: An evidential network approach ». *Reliability Engineering & System Safety,* vol. 162, p. 111-121.