HAL
open science

Méthode d'estimation des valeurs extrêmes des distributions de contraintes induites et de seuils de susceptibilité dans le cadre des études de durcissement et de vulnérabilité aux menaces électromagnétiques intentionnelles

Thomas Houret

To cite this version:

Thomas Houret. Méthode d'estimation des valeurs extrêmes des distributions de contraintes induites et de seuils de susceptibilité dans le cadre des études de durcissement et de vulnérabilité aux menaces électromagnétiques intentionnelles. Electromagnétisme. INSA de Rennes, 2019. Français. NNT : 2019ISAR0011 . tel-02521324

HAL Id: tel-02521324

https://theses.hal.science/tel-02521324

Submitted on 27 Mar 2020

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

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

THESE DE DOCTORAT DE

L'INSA RENNES

COMUE UNIVERSITE BRETAGNE LOIRE

ECOLE DOCTORALE N° 601
Mathématiques et Sciences et Technologies de l'Information et de la Communication
Spécialité : Electronique et Télécommunications

Par

Thomas HOURET

Méthode d'estimation des valeurs extrêmes des distributions de contraintes induites et de seuils de susceptibilité dans le cadre des études de durcissement et de vulnérabilité aux menaces électromagnétiques intentionnelles

Thèse présentée et soutenue à Rennes le 19 décembre 2019
Unité de recherche : CEA Gramat, IETR
N° d'ordre : 19ISAR 37 / D19 - 37

Rapporteurs avant soutenance :

Lionel PICHON (Président du Jury)
Directeur de Recherche CNRS, GeePs, Gif-sur-Yvette

Sébastien LALLECHERE
Maitre de conférences, HDR, Université Clermont Auvergne, Clermont-Ferrand

Composition du Jury :

Elodie RICHALOT (Examinatrice)
Professeur des universités, Université Paris Est Marne-la-Vallée, Champs-sur-Marne

Alain REINEIX (Examinateur)
Directeur de recherche CNRS, XLIM, Limoges

Philippe POULIGUEN (Examinateur)
Ingénieur RDS « ondes acoustiques et radioélectriques », HDR, Agence de l'innovation de défense, DGA, Paris

Philippe BESNIER (Directeur de thèse)
Directeur de recherche CNRS, IETR, Rennes

Stéphane VAUCHAMP (Co-encadrant de thèse)
Ingénieur en électromagnétisme, CEA/DAM, Gramat

Invités :

Michel BOURZEIX
Assistant scientifique, CEA/DAM, Gramat

Elodie BATAILLER
Responsable expertises en simulation numérique, DGA-TA, Toulouse

Méthode d'estimation des valeurs extrêmes des distributions de contraintes induites et de seuils de susceptibilité dans le cadre des études de durcissement et de vulnérabilité aux menaces électromagnétiques intentionnelles

Thomas Houret

Document protégé par les droits d'auteur

REMERCIEMENTS

« Se tenir sur les épaules des géants et voir plus loin, voir dans l'invisible, à travers l'espace et à travers le temps » Jean Claude Ameisen.

Cette thèse n'est pas le fruit d'une performance individuelle. J'ai effectivement l'impression de me tenir sur les épaules de « géants » qui m'ont permis d'aboutir à cette très modeste contribution qui, je l'espère, nous permettra de « voir plus loin ». Cette page est l'occasion de remercier tous ceux qui ont participé à cette contribution.

L'argent étant le nerf de la guerre, commençons par le **CEA** pour avoir financé ma thèse, c'est-à-dire avoir assuré mon salaire pendant trois ans.

Je remercie ensuite mes deux rapporteurs, L. **Pichon** et S. **Lalléchère**, sans qui cette page ne serait pas écrite. Je tiens aussi à remercier E. **Richalot**, A. **Reineix**, P. **Pouliguen**, M. **Bourzeix** et E. **Batailler**, pour avoir relu mon manuscrit avec intérêt et d'avoir constitué mon jury le jour J. Malgré les grèves et les problèmes techniques, on a fini par y arriver (merci **Laurent** et **Jérémy** !).

Evidemment, je remercie **Philippe**, mon directeur de thèse. Directeur de recherche au CNRS, ton excellence académique a été un atout précieux. Même après 17 thèses suivies, ta motivation est toujours intacte. Ton implication dans ma thèse a été très conséquente et sans égale, n'en déplaise aux autres doctorants et leurs encadrants. D'ailleurs, cette thèse est autant le fruit de ton travail que du mien. Tes qualités humaines, notamment ton humour, ont été aussi un facteur déterminant. Bref, ces trois années à travailler à tes cotés ont été tout simplement fantastiques !

Ensuite, vient forcement **Stéphane**, chercheur au CEA de Gramat, mon co-encadrant de thèse. Tu as eu le courage d'accepter de m'encadrer, ton expertise n'étant pas les statistiques. Mais tu n'en avais pas besoin pour me comprendre. Tu m'as même dit un jour que finalement, ce n'était pas compliqué ce que je faisais. Tu m'as aussi apporté une aide méthodologique et rédactionnelle, fondamentale à ma réussite. Enfin, tu m'as grandement facilité les démarches administratives, ce qui au CEA n'est pas inutile (cf EMC Europe Barcelone).

Je voudrais aussi remercier ceux qui ont participé indirectement à la réussite de cette thèse. Merci à **Yvan**, **Kais**, **Otman** et **Odile**, pour nos pauses café-clope qui ont certes abimé mes poumons, mais largement égayé mes journées à l'IETR. Je n'oublie pas mes collègues sans nicotine : **Abdou**, **Jessen**, **Florent**, **Ludivine**, **Faissal**, **Yi, Thomas**, **Clovis**.

Je remercie particulièrement **Abdivall**, pour nos nombreuses escapades en Bretagne, à Paris, Amsterdam et Barcelone. Désolé de t'avoir conseillé ces pâtes carbo dans le XVème. Tu t'es vengé quand on a rendu visite à Chateaubriand. A peu de chose près, j'appelais la SNSM, histoire d'être la risée de tous les corsaires de la ville !

Merci aussi à **Basil** d'avoir régulièrement vidé ma poubelle et fais tes besoins sur ma moquette. J'en profite pour remercier **Quentin** et **Gladys** pour l'avoir gardé quand j'étais en vadrouille aux frais du contribuable, aux quatre coins du monde.

Et pour finir, un grand merci à toute ma famille (**Yves**, **Alix**, **Fanny**, **Caroline**) pour ces rares moments passés ensemble. Certains ont été moins drôles que d'autres (août 2018 à Monaco), mais la plupart m'ont permis de revenir au top.

RESUME

Une Interférence Electromagnétique Intentionnelle (IEMI) agressant un équipement électronique peut provoquer une défaillance de ce dernier. L'étude des effets d'une IEMI commence par l'évaluation du risque de défaillance afin de mettre en place les protections adéquates. Malheureusement, une prédiction déterministe d'une défaillance est impossible car les caractéristiques de l'équipement et de l'agression sont très incertaines. La stratégie proposée dans cette thèse consiste à modéliser la contrainte générée par l'agression, ainsi que la susceptibilité de l'équipement, comme des variables aléatoires. Ensuite, trois étapes sont nécessaires. La première concerne l'estimation de la distribution probabiliste de la variable aléatoire de susceptibilité, la seconde porte sur celle de la contrainte, pour enfin, dans une troisième étape, en déduire la probabilité de défaillance.

Pour la première étape, nous utilisons des méthodes d'inférence statistique sur un petit échantillon de seuils de susceptibilité mesurés. Nous comparons deux types d'inférence paramétrique : bayésienne et celle du maximum de vraisemblance. Nous concluons qu'une approche pertinente pour une analyse de risque consiste à utiliser les intervalles de confiance, ou de crédibilité des estimations des paramètres, pour encadrer la probabilité de défaillance, quelle que soit la méthode d'inférence choisie.

Pour la deuxième étape, nous explorons les techniques de recherche de valeurs extrêmes tout en réduisant le nombre de simulations nécessaires. En particulier, nous proposons la technique de la stratification contrôlée par un métamodèle de krigeage. Nous montrons que cette technique améliore drastiquement les performances par rapport à l'approche classique (simulation Monte Carlo). De plus, nous proposons une implémentation particulière de cette technique afin de maîtriser le coût de calcul.

Enfin, la troisième étape est la plus simple une fois les deux premières franchies puisque, par définition, une défaillance survient lorsque la contrainte est supérieure à la susceptibilité. A partir d'un cas test final comportant la simulation de l'agression d'un équipement et de données de susceptibilité de cet équipement, nous calculons un encadrement de la probabilité de défaillance en recourant aux méthodes développées pendant cette thèse. En particulier, nous montrons que l'utilisation conjointe de la stratification contrôlée par le krigeage, et de l'inférence de la distribution de la susceptibilité, permet effectivement d'encadrer l'estimation de la vraie probabilité de défaillance.

ABSTRACT

Intentional ElectroMagnetic Interference (IEMI) can cause equipment failure. The study of the effects of an IEMI begins with an assessment of the risk of equipment failure in order to implement appropriate protections, if required. Unfortunately, a deterministic prediction of a failure is impossible because both characteristics of equipment and of the aggression are very uncertain. In this Ph. D. thesis, the proposed strategy consists of modelling the stress generated by the aggression, as well as the susceptibility of the equipment, as random variables. Then, three steps are necessary: The first step deals with the estimation of the probability distribution of the susceptibility random variable. The second step deals with the similar estimation for the constraint / stress random variable. Eventually, the third step concerns the calculation of the probability of failure.

For the first step, we use statistical inference methods on a small sample of measured susceptibility thresholds. We compare two types of parametric inference: bayesian and maximum likelihood. We conclude that a relevant approach for a risk analysis is to use the confidence or credibility intervals of parameter estimates to frame the probability of failure, regardless of the inference method chosen.

For the second step we investigate extreme value exploration techniques while reducing the number of simulations required. In particular, we propose the technique of Controlled Stratification by Kriging. We show that this technique drastically improves performance compared to the classic approach (Monte Carlo simulation). In addition, we propose a particular implementation of this technique in order to control the calculation effort.

Finally, the third step is the simplest once the first two steps have been completed since, by definition, a failure occurs when the stress is greater than the susceptibility. With the help of a final test case comprising the simulation of an electromagnetic aggression on a piece of equipment, we use the method developed in our work to estimate the frame of the probability of failure. More specifically, we show that the combined use of controlled stratification by kriging and inference of susceptibility distribution, allows framing the estimated true value of the probability of failure.

SOMMAIRE

Chapitre I. Introduction générale 11

Chapitre II. Analyse probabiliste de la susceptibilité d'équipements électroniques 17

I. Introduction 17

II. Etat de l'art 17

III. Estimation du type de distribution 27

IV. Estimation des paramètres de la distribution 31

V. Encadrement probabiliste du seuil de susceptibilité 34

VI. Applications CEM 39

VII. Conclusion et pistes de réflexion 46

Chapitre III. Recherche de valeurs extrêmes de contrainte par krigeage et stratification contrôlée 49

I. Introduction 49

II. Etat de l'art 50

III. Modèles utilisés 61

IV. Comparaison des performances des métamodèles 65

V. Intérêt de la Stratification Contrôlée par rapport au Krigeage seul 70

VI. Gestion du budget de simulation lors de l'estimation d'un quantile extrême par la méthode K-SC 77

VII. Conclusion et pistes de réflexion 93

Chapitre IV. Estimation de la probabilité de défaillance d'équipements électroniques en cas d'agression électromagnétique 95

I. Introduction 95

II. Présentation du cas d'étude 96

III. Obtention d'une probabilité de défaillance de référence 99

IV. Estimation de la probabilité de défaillance 101

V. Conclusion et piste de réflexion 103

Chapitre V. Conclusion générale 105

Références 107

Annexes 115

I. Annexes du Chapitre II 115

II. Annexes du Chapitre III 117

III. Listes des Figures et Tableaux 119

IV. Liste des publications 126

VI. Table des matières 127

Tableau 1: Principaux acronymes, et conventions de notations mathématiques utilisées.

Acronyme	**Description**	**Exemple de notation**	**Description**
CEM	Compatibilité Electromagnétique	X	Une variable aléatoire
IEMI	Interférence Electromagnétique Intentionnelle	x	Une valeur de réalisation de la variable aléatoire X
IB	Inférence Bayésienne	$\mathbb{P}\{A\}$	Probability qu'un événement, ici A, se réalise
IMV	Inférence selon le Maximum de Vraisemblance	$\boldsymbol{S}$	Un vecteur
N	Loi Normale	S_1	La première composante du vecteur $\boldsymbol{S}$
LN	Loi Log-normale	μ_X	Moyenne de la variable X
W	Loi Weibull	$E[X]$	Espérance de la variable X
AD	Test d'Anderson-Darling	σ_X	Ecart type de la variable X
BIC	*Bayesian Inference Criterion*	$V[X]$	Variance de la variable X
DOE	*Design Of Experiment* (plan d'expérience)	$f_X(x)$	*pdf* de la variable X évaluée en x
CV	Coefficient de Variation	$F_X(x)$	*cdf* de la variable X évaluée en x
SC	Stratification Contrôlée	$\mathbb{I}_{condition}$	Fonction indicatrice Par exemple : $\mathbb{I}_{X<1} = 1$ si $X < 1$, 0 sinon
SVM	*Support Vector Machine*	$\wedge, \vee$	Et, ou logiques
NN	*Neural Network* (Réseau de neurones)	y_α	Un quantile de Y dont la probabilité associée est α
PCE	*Polynomial Chaos Expansion*	quantile α	Un quantile, sans nom explicite, associé à la probabilité α
KRI	krigeage	$\hat{\theta}$	Un estimateur de θ
SM	*Surrogate Model* (métamodèle)	$\mid$	Sachant, étant donnée
MCS	*Monte Carlo Simulation*	$\lfloor X \rfloor$	Partie entière de X (*floor*)
EE	Estimation Empirique (fournie par MCS)	*mod*	modulo
pdf	*Probability density function* (fonction de densité)	*1.3*	Le point comme séparateur décimale
cdf	*Cumulative density function* (fonction de densité cumulée)		
i.e.	*Id est* (c'est-à-dire)		
e.g.	*Exempli gratia* (par exemple)		
etc.	*Et cetera*		

Tableau 2: Variables spécifiques au chapitre II.

Variable	Description
$[s_{low}, s_{upp}]$	Bornes encadrant s_α
n	Taille de l'échantillon
e	Numéro d'équipement à tester ($e=1 \ldots n$)
S_e	Seuil de susceptibilité mesuré pour l'équipement e
nt_e	Dernier niveau appliquée
δ_e	Résolution des données mesurés
$\boldsymbol{\theta}$	Vecteur paramètre de la *pdf/cdf.* *Cas N :* $\boldsymbol{\theta} = [\mu, \sigma^2]$, cas LN: $\boldsymbol{\theta} = [a, b^2]$, cas W : $\boldsymbol{\theta} = [\gamma, \beta]$
f_{post}	*pdf* a posteriori de $\boldsymbol{\theta}$
f_{prior}	*pdf* a priori de $\boldsymbol{\theta}$
f_{Like}	Distribution de vraisemblance
$\boldsymbol{Data}$	Vecteur contenant l'échantillon des seuils mesurés
$p_f(s)$	Probabilité de défaillance évaluée en s
$[L(s), U(s)]$	Bornes encadrant $p_f(s)$

Tableau 3: Variables spécifiques au chapitre III.

Variable	Description	Variable	Description
X	Variable aléatoire d'entrée	μ_{qind}	Moyenne des MC quantiles estimés par SM_{ind}
Y	Variable aléatoire de sortie	$q_{ind,mc}$	Quantile estimé par SM_{ind} à la mceme boucle
X_{max}	DOE (entrée) de SM_{max}	σ_{qind}	Ecart type des MC quantiles estimés par SM_{ind}
Y_{max}	DOE (sortie) de SM_{max}	N_j	Budget à chaque $j^{ème}$ strate
Xcs	DOE (entrée) pour la SC	n	Nombre d'appel au modèle coûteux
SM_{ind}	SM construit avec un DOE de taille S_{ind}	n_s	Nombre de strates (stratification contrôlée)
S_{SM}	Budget dédié au SM	n_{ext}	Nombre de valeurs extrêmes
$Smax$	Taille de DOE max pour le SM	N_{min}	Taille de DOE minimale pour observer au moins next valeurs extrêmes
$S_{SM,add}$	Budget à ajouter à S_{SM}	$\boldsymbol{A}$	Vecteur de probabilités définissant les strates
α	Probabilité du quantile visé	$Y_i^{(j)}$	ième réalisation de sortie à partir de la $j^{ème}$ strate

Tableau 4: Variables spécifiques au chapitre IV.

Variable	Description
n_K	Taille de l'échantillon de contrainte dédiée à la construction du krigeage
n_{SC}	Taille d'échantillon générée par la stratification controlée
n_C	Budget totale de simulation pour estimer la contrainte ($n_C = n_{SC} + n_K$)
n_S	Taille de l'échantillon de seuils de susceptibilité

CHAPITRE I. INTRODUCTION GENERALE

I. Contexte

La Compatibilité Electromagnétique (CEM) est un domaine d'étude visant à assurer que plusieurs équipements électroniques d'un même environnement, fonctionnent normalement. Ainsi, les équipements doivent être immunisés face à de possibles émissions électromagnétiques de la part d'équipements voisins. Ces émissions constituent une source d'interférences électromagnétiques pouvant perturber le fonctionnement d'un équipement. Il s'agit, en générale, de sources d'interférences involontaires.

Cependant, dans le cadre de la guerre électronique, une source peut être intentionnelle, avec pour objectif de provoquer une interférence dans le fonctionnement d'un équipement victime. Les Interférences Electromagnétiques Intentionnelles (IEMIs) constituent une composante majeure de la guerre électronique. En effet, une protection efficace des applications logicielles n'est d'aucune utilité si celle de l'architecture matérielle est délaissée. La menace des IEMIs concerne tous les secteurs industriels : celui de l'énergie [1], de l'automobile [2], de l'aéronautique [3] et des drones (Figure 1), du ferroviaire [4], du spatial [5], ou encore celui des objets connectés (IoT) [6]. Des cas d'agression par IEMI ont déjà été répertoriés et analysés [7].

Figure 1 : Le NEROD F5 de MC2 technologies. Cette arme permet de générer une IEMI ciblée afin de forcer l'atterrissage d'un drone en toute sécurité. Photo datant du défilé militaire du 14 juillet 2019 à Paris.

Le Commissariat à l'Energie Atomique et aux énergies alternatives (CEA) de Gramat, qui a initié les travaux de cette thèse, est spécialisé dans l'étude des effets des armes, notamment dans le domaine électromagnétique. Deux types d'IEMI y sont étudiés : les Armes à Energie Dirigée (AED) et les Impulsions Electromagnétiques d'origine Nucléaire (IEMN).

Les AED, comme leur nom l'indique, concentrent de l'énergie électromagnétique précisément sur une cible. Un large spectre de fréquences peut être utilisé par les AED, allant des fréquences radio [8][9] jusqu'aux fréquences laser [10].

L'IEMN est une impulsion de fort niveau générée lors de l'explosion d'une bombe nucléaire à haute altitude, qui se propage vers le sol avec une très grande couverture sur un territoire. Tous les équipements électroniques d'un pays risquent alors une panne, voire même la destruction. Les effets matériels d'une attaque de type IEMN constituent toujours un axe de recherche majeur, d'importance stratégique [11]. Même après la guerre froide, de telles attaques sont toujours d'actualité [12].

II. Problématique et objectif

L'étude des effets d'une agression par IEMI commence par l'évaluation du risque de défaillance des équipements afin de mettre en place les mesures de protection adéquates. Depuis 1864, il est possible de décrire mathématiquement le comportement des ondes électromagnétiques en résolvant quatre équations différentielles, communément appelées équations de Maxwell [13]. Les progrès informatiques permettent aujourd'hui de résoudre ces équations numériquement avec une grande précision. Il existe actuellement de nombreux logiciels, commerciaux ou internes, qui offrent la possibilité de résoudre ces équations quel que soit le cas électromagnétique étudié, aussi complexe soit-il du point de vue de la géométrie ou des caractéristiques intrinsèques des matériaux. Il serait donc possible en théorie de prédire exactement l'intensité d'un courant causé par une IEMI, dans quelque zone que ce soit à l'intérieur de l'équipement victime.

Malheureusement, même en disposant de modèles déterministes, une telle prédiction est impossible car les caractéristiques de l'équipement et de l'agression sont très incertaines. L'origine des incertitudes est toutefois connue : il s'agit d'incertitudes épistémiques [14] liées à une mauvaise connaissance, qu'il faut distinguer d'incertitudes naturellement aléatoires. Par exemple, la désintégration radioactive est un phénomène considéré comme fondamentalement aléatoire contrairement à l'électrodynamique classique qui décrit des phénomènes physiques déterministes. Cependant, cette distinction n'aura, en ce qui concerne le problème posé plus loin, aucune influence sur les méthodes de traitement de ces incertitudes. Ces incertitudes affectent à la fois la contrainte et la susceptibilité, termes que nous allons à présent définir.

Une contrainte est une grandeur physique (courant, tension, champ électrique ou magnétique, etc.), résultante de l'agression IEMI suivant un chemin de couplage, qui agresse un équipement à son ou ses entrée(s). De multiples incertitudes interviennent lorsqu'il s'agit de décrire la contrainte, comme par exemple : l'intensité et l'angle d'incidence de l'IEMI, l'environnement externe de l'équipement, etc.

Le seuil de susceptibilité d'un équipement représente le niveau d'une grandeur physique à son entrée responsable d'une défaillance. Ce niveau peut être par exemple un courant sur une entrée de l'équipement, au-delà duquel une défaillance survient. Cependant, ce seuil de susceptibilité n'est pas une constante du fait des incertitudes concernant l'équipement. Ces incertitudes viennent par exemple d'un plan de design non fourni, d'un processus de fabrication impliquant une variabilité entre plusieurs équipements pourtant issus de la même série, de la tolérance des composants électroniques, etc.

Lors de l'évaluation du risque de défaillance d'un équipement soumis à une IEMI, la problématique est donc la suivante. Les multiples sources d'incertitudes épistémiques impliquent que la contrainte devienne incertaine, tout comme la susceptibilité. De plus, nous supposons que les équipements sont suffisamment bien conçus pour n'être susceptibles qu'à des valeurs de contraintes élevées, i.e. des valeurs extrêmes. Par conséquent, le risque de défaillance est faible. Face aux incertitudes, quelle approche peut permettre d'évaluer convenablement ce risque, surtout si celui-ci est faible ?

Cette problématique est illustrée par la Figure 2. Nous avons représenté trois valeurs possibles de contraintes (c_1, c_2 et c_3) agressant trois équipements. A cause des incertitudes, il est impossible de prédire avec certitude laquelle se réalisera. De plus, les trois équipements ne sont pas sensibles à la même valeur de contrainte les agressant. En effet, bien qu'issus du même modèle, leurs seuils de susceptibilité respectifs peuvent aussi prendre plusieurs valeurs possibles. Ces trois équipements sont susceptibles à trois valeurs de seuils : s_1, s_2 et s_3. L'unité de la contrainte et de la susceptibilité est ici volontairement omise. Il pourrait s'agir d'Ampères, de Volts ou toute autre unité physique, tant que celle-ci soit commune à la contrainte et la susceptibilité. La susceptibilité du premier équipement est s_1. Or celui-ci est agressé par une valeur de contrainte c_1, supérieure à s_1. Par conséquent, l'équipement ne peut plus fonctionner correctement. Selon le critère de défaillance choisi, il peut s'agir d'un simple

dysfonctionnement temporaire, voire d'une destruction définitive. Pour les deux autres équipements, la contrainte est inférieure à la susceptibilité, ce qui n'engendre pas de défaillances. Au regard de la Figure 2, il apparait que plus la probabilité de défaillance est faible, plus la configuration s_1, c_1 est rare. En imaginant que les modèles de simulation soient disponibles, tant pour la contrainte que pour la susceptibilité, une détermination empirique du risque suppose la réalisation d'un très grand nombre de couples de simulations de valeurs de s_i, c_i. Cependant, les modèles de simulation de susceptibilité ne sont en général pas disponibles, ce qui constitue un verrou supplémentaire.

L'objectif de cette thèse est de proposer une méthodologie pour quantifier le risque de défaillance d'équipements en cas d'agression de type IEMI. Plus précisément, nous cherchons à quantifier le risque qu'un équipement, dont la susceptibilité est incertaine, soit défaillant lorsqu'il est agressé par une contrainte, elle aussi incertaine. Le fait que le risque soit faible constitue une difficulté supplémentaire.

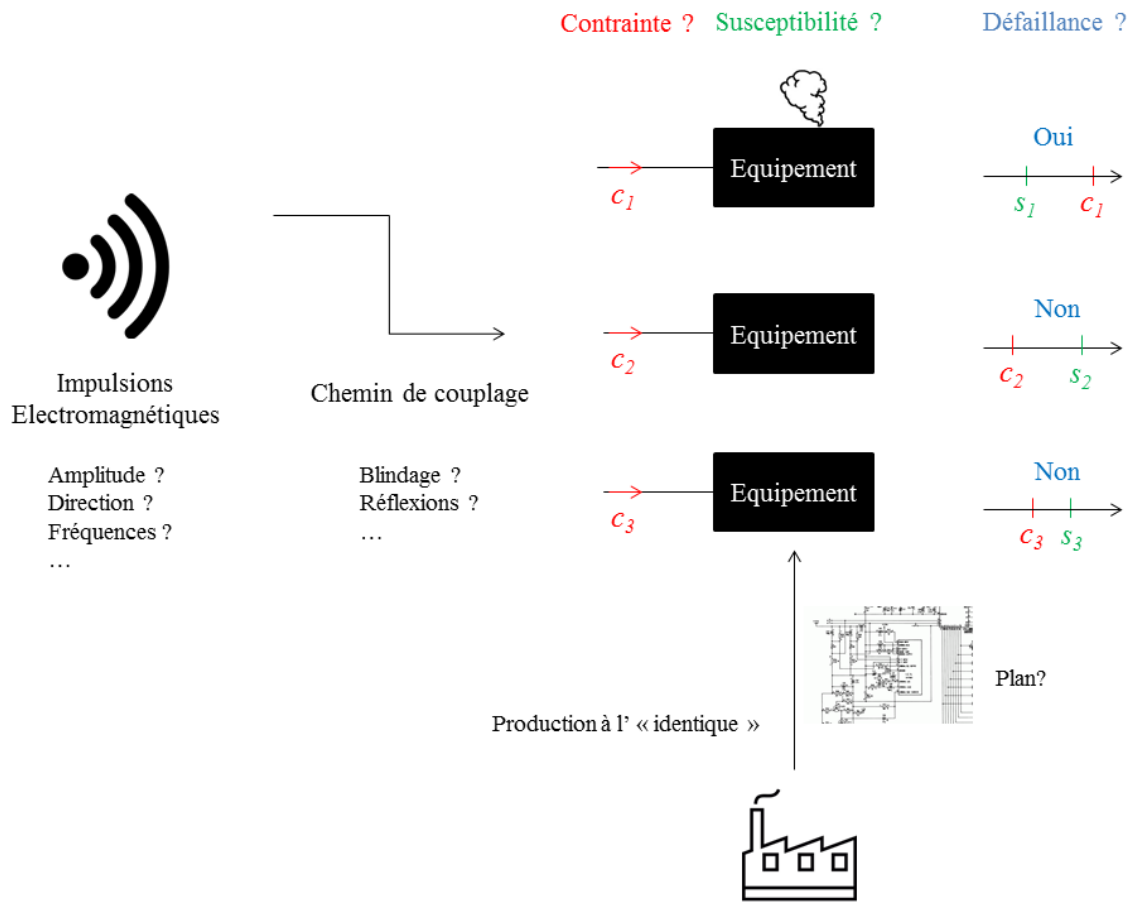

Figure 2 : Schéma résumant la problématique d'une étude de risque en cas d'IEMI : les incertitudes épistémiques (termes associés aux points d'interrogation) impliquent une contrainte (rouge) et une susceptibilité (verte) incertaines.

III. Positionnement de la thèse

Notre stratégie commence par définir la contrainte ainsi que la susceptibilité comme des variables aléatoires. Au sens des phénomènes physiques étudiés dans cette thèse, définir la contrainte comme une variable aléatoire est bien adapté à la représentation de l'agression électromagnétique qui dépend elle-même de nombreux paramètres. De même, l'absence d'un modèle de susceptibilité, justifie le choix d'une représentation de cette dernière par une variable aléatoire.

L'étude de la susceptibilité est réalisée à partir de mesures. En effet, à l'heure actuelle, il est

impossible de simuler le fonctionnement d'équipements complexes compte tenu de l'indisponibilité de modèles complets de composants internes (notamment des circuits intégrés en dehors de leurs régimes nominaux de fonctionnement) et des incertitudes associées. En métrologie, une approche stochastique pour modéliser les incertitudes est déjà unanimement acceptée [15][16]. D'ailleurs, des guides existent pour l'utilisateur [17] comme pour la rédaction des normes CEM [18]. Cependant, il existe encore de mauvaises pratiques, même à des fins normatives, négligeant la variabilité aléatoire inhérente au processus de mesure. Cette variabilité peut être pourtant considérable : jusqu'à 10 *dB* d'incertitudes pour des systèmes complexes [19]. De plus, même lorsqu'une approche statistique rigoureuse est mise en place, son potentiel n'est pas exploité au maximum.

En simulation, le recours à l'approche stochastique est beaucoup plus récent [20]. Cela vient en grande partie du fait que cette approche n'est pas naturelle, puisque la propagation électromagnétique est fondamentalement déterministe. De plus, les moyens de simulation (e.g. HFSS, CST, FEKO, etc.) sont aujourd'hui très précis, à condition que l'on connaisse parfaitement les paramètres du problème. Or, cette connaissance est souvent très imparfaite, ce qui est rarement pris en compte, même si récemment différentes approches ont été proposées sous forme de métamodèles ou « surrogate models » (SMs) selon la terminologie anglo-saxonne.

Le risque de défaillance est quantifié par la probabilité que la variable aléatoire de la contrainte prenne une valeur supérieure à celle de la susceptibilité. L'estimation de cette probabilité consiste à procéder en trois étapes résumées dans le schéma de la Figure 3, qui sont : l'estimation de la distribution de la variable aléatoire de susceptibilité, puis celle de la contrainte, et enfin la déduction de la probabilité de défaillance.

La première étape consiste à estimer la distribution de la susceptibilité de l'équipement à partir d'un échantillon de mesures. Les campagnes de mesures prennent souvent du temps, le nombre d'équipements testés est ainsi limité. L'objectif de cette étape est donc d'estimer la distribution de susceptibilité (courbe verte de la Figure 3) à partir d'un nombre réduit de mesures. La stratégie proposée dans cette thèse consiste à utiliser des méthodes d'inférence statistique.

La deuxième étape consiste à estimer la distribution de la contrainte à partir d'un nombre limité d'appels à un modèle déterministe basé sur la résolution des équations de Maxwell. Il s'agit donc d'estimer la courbe rouge de la Figure 3. La résolution étant coûteuse en temps de calcul, il est impossible d'estimer la distribution à partir d'un très grand échantillon de simulations. De plus, nous recherchons des valeurs extrêmes de contraintes. Ces valeurs de contraintes sont dites extrêmes, tant par leur haut niveau que par leur rareté. Autrement dit, leur probabilité d'occurrence est faible mais leur réalisation est potentiellement dangereuse. C'est pourquoi nous avons exploré les techniques de recherche de valeurs extrêmes tout en réduisant le nombre de simulations nécessaires. Bien que simple et robuste, la variance des estimateurs fournie par une approche empirique, comme la simulation de Monte Carlo (MCS), est importante. C'est pourquoi nous avons exploré des techniques dites de réduction de variance.

Enfin, la troisième étape consiste à calculer la probabilité de défaillance connaissant la distribution des valeurs possibles de la susceptibilité et de la contrainte. Lorsque la contrainte est supérieure à la susceptibilité, une défaillance survient. Il s'agit donc d'une zone à risque (en bleu dans la Figure 3).

En résumé, cette thèse s'inscrit délibérément à la suite des travaux récents consacrés aux approches statistiques en CEM. Ces travaux sont essentiellement basés sur la famille des outils de la mathématique statistique et concernent l'obtention de SMs ou les techniques de réduction de la variance. Ils font notamment suite aux travaux de thèse de M. Mourad Larbi (CEA / IETR) [21] qui concernaient uniquement l'étude de la contrainte. Nous étendons ces travaux à cette approche globale d'analyse de risque en apportant les contributions majeures soulignées dans la section suivante.

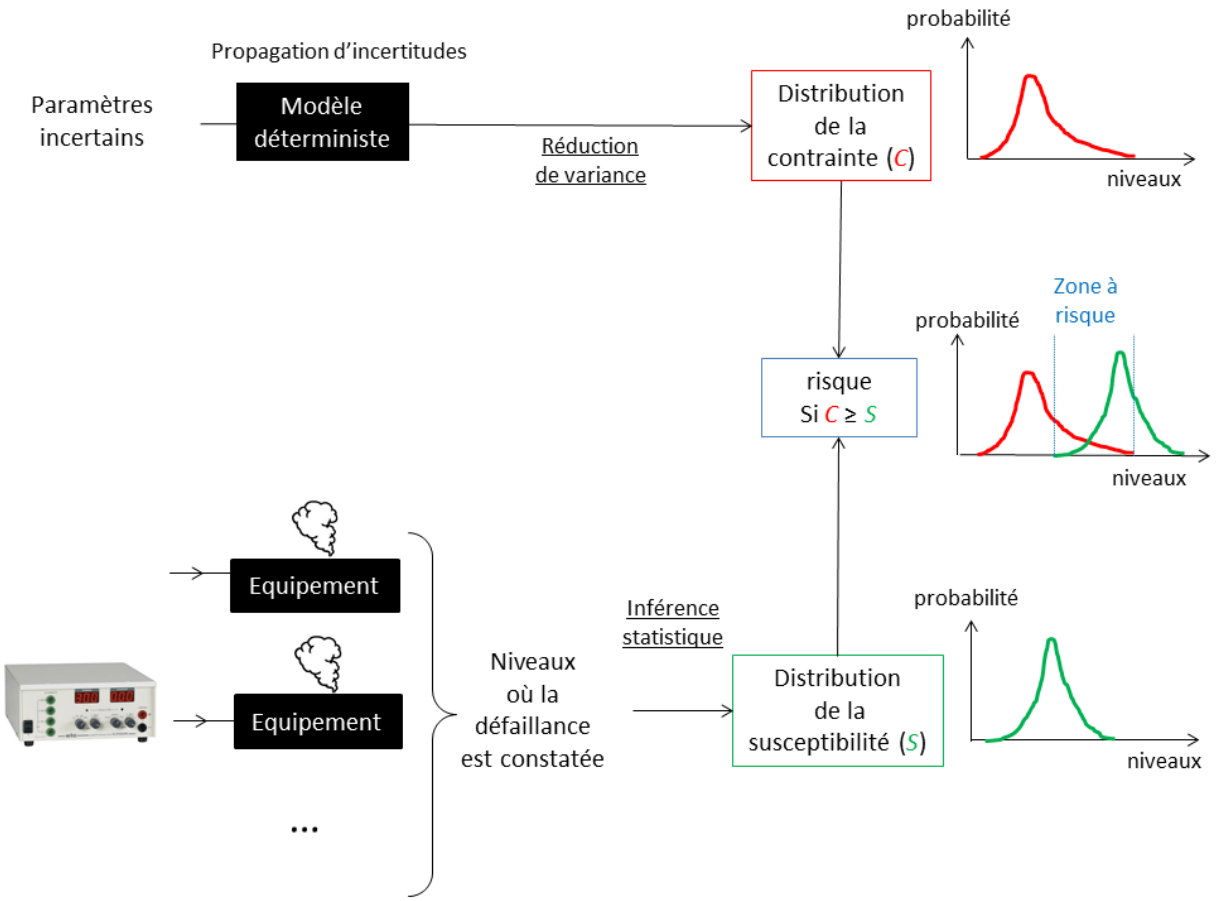

Figure 3 : Schéma résumant notre stratégie pour évaluer la probabilité de défaillance d'équipements soumis à une agression de type IEMI. Les encadrés (bleu, vert et rouge) correspondent aux trois étapes de notre stratégie. Les termes soulignés correspondent aux solutions techniques mises en place.

IV. Organisation du manuscrit

Le manuscrit est organisé en trois chapitres, hormis l'introduction et la conclusion générale (respectivement chapitres I et V) :

Le chapitre II est exclusivement dédié à l'estimation de distribution de la susceptibilité. Un état de l'art permet de décrire les différentes méthodes d'inférence statistique. Au sein de ce chapitre, l'apport majeur de cette thèse (section V) est de rappeler à la communauté CEM qu'il est possible, et même souhaitable, d'encadrer la distribution de susceptibilité avec un intervalle de confiance. L'incertitude concernant les seuils de susceptibilité mesurés est donc quantifiable en fonction du nombre d'équipements testés.

Le chapitre III se concentre sur l'estimation de la contrainte en appliquant une technique de réduction de variance appelée stratification contrôlée. Un état de l'art permet de rappeler les différentes méthodes d'estimation de quantile et en particulier l'utilisation de SMs et de la stratification contrôlée. Au sein de ce chapitre, l'apport majeur de cette thèse consiste à associer la technique de la stratification contrôlée avec du krigeage (section V), puis de proposer un algorithme d'implémentation particulier permettant de maîtriser le budget de simulation (section VI).

Dans le chapitre IV, les estimations de la contrainte et de la susceptibilité sont réunies pour traiter un cas d'étude pratique. Ce chapitre apporte une contribution supplémentaire concernant l'étude de la contrainte, en proposant une stratégie de stratification adaptée à l'estimation de la probabilité de défaillance (section IV).

CHAPITRE II. ANALYSE PROBABILISTE DE LA SUSCEPTIBILITE D'EQUIPEMENTS ELECTRONIQUES

I. Introduction

L'évaluation du risque de défaillance des équipements est requise pour mettre en place des protections électromagnétiques, notamment pour faire face aux IEMI pour des équipements dont le rôle est critique. Les équipements électroniques sont différents les uns des autres et complexes. Bien qu'il ne soit pas impossible d'établir des modèles pour certains composants les constituant, il est très difficile de modéliser l'ensemble de l'équipement pour des simulations de susceptibilité. De plus, la géométrie, les caractéristiques électriques de chaque composant ainsi que les procédés de fabrication sont entachés d'incertitudes. Par conséquent, le seuil de susceptibilité de plusieurs équipements varie aléatoirement même s'ils sont tous issus d'un même modèle, i.e. ayant le même schéma de conception et remplissant le même rôle. La susceptibilité d'un équipement est donc mieux décrite par une distribution probabiliste de son seuil de susceptibilité qui est alors défini comme une variable aléatoire. L'objectif de l'analyse de susceptibilité consiste donc à estimer la distribution probabiliste du seuil de susceptibilité de l'équipement considéré.

Ce chapitre est organisé en 5 sections, hormis l'introduction (section I) et la conclusion et pistes de réflexions (section VII).

Dans la section II, nous rappelons et détaillons l'état de l'art concernant la démarche expérimentale de mesure de seuils de susceptibilité puis du traitement de ces données par inférence statistique. La sous-section 2 rappelle notamment que les inférences paramétriques sont basées sur le choix d'une distribution sous-jacente. L'expression analytique de la fonction de répartition étant supposée connue, les inférences paramétriques cherchent à estimer les paramètres $\boldsymbol{\theta}$ de la fonction de répartition. Il y a donc deux sources d'erreurs potentielles : le choix de l'expression de la fonction de répartition puis de l'estimation de $\boldsymbol{\theta}$.

Nous avons procédé à deux études théoriques dans lesquelles les caractéristiques de la distribution de susceptibilité sont connues a priori. Dans la section III, nous évaluons l'erreur du choix du type de loi (donnée par sa fonction de répartition) sans analyser $\boldsymbol{\theta}$. Dans la section IV, nous nous concentrons sur les estimations de $\boldsymbol{\theta}$ sachant l'expression de la fonction de répartition. Ces deux sections (III et IV) utilisent uniquement un estimateur scalaire de $\boldsymbol{\theta}$, ce qui ne permet pas d'exploiter tout le potentiel des méthodes d'inférence.

Puis dans la section V, nous encadrons la probabilité de défaillance des seuils de susceptibilité. Cette section constitue l'apport majeur de cette thèse pour la communauté CEM. En effet, dans cette section nous exploitons l'ensemble des valeurs possible de l'estimateur de $\boldsymbol{\theta}$. De plus, l'identification de la loi de probabilité sous-jacente de la susceptibilité et l'estimation des paramètres associés sont regroupées.

Enfin, dans la section VI nous appliquons les méthodes d'inférences sur 2 cas tests où les caractéristiques de la loi sont inconnues a priori. Le premier cas test concerne la susceptibilité d'équipements virtuels tandis que le second traite des mesures effectuées sur des alimentations.

II. Etat de l'art

1. Démarche expérimentale des mesures de susceptibilité

Nitsch et al. ont soumis à une IEMI une gamme de circuits à base de microprocesseurs ou de microcontrôleurs ayant des comportements différents selon leur technologie interne [22]. La nature

aléatoire des câbles ou des pistes de circuits imprimés a été abordée d'un point de vue théorique et statistique dans [23] et [24], les composants actifs n'ayant pas été considérés. Cependant, une analyse topologique prévisionnelle de haut niveau a été proposée dans [25]. Notons également que [26] fournit un cadre général pour effectuer une analyse des risques par arbre de défaillances dont le calcul nécessite une connaissance des fonctions de distribution des paramètres impliqués. Il ressort de ces publications qu'une approche expérimentale reste probablement toujours la plus accessible. Il semble en effet relativement illusoire de traiter de la susceptibilité par le seul moyen de la simulation numérique des mécanismes entrainant la susceptibilité. Une approche expérimentale se limite cependant à un petit nombre d'équipements. En effet, en raison de contraintes de temps et de coût, une dizaine d'unités seulement sont généralement disponibles pour les tests. Nous proposons donc d'estimer au mieux la distribution probabiliste du seuil de susceptibilité à partir d'un jeu de données, nécessairement limité.

Cette approche expérimentale suppose que chaque équipement soit caractérisé par un seuil de susceptibilité S_e unique issu d'une distribution f_S commune à l'ensemble des équipements du même modèle. La Figure 4 illustre la procédure mise en place en pratique lors des campagnes de test de susceptibilité. Un échantillon de *n* équipements, généralement de petite taille, est mis à disposition pour les tests. Un indice, e, désigne chacun des équipements (*e=1…n*). Chaque copie d'équipement est testée l'une après l'autre. On suppose que le test d'un équipement n'influe pas sur celui du suivant, i.e. les seuils mesurés sont indépendants. Le test consiste à augmenter le niveau de contrainte (e.g. l'intensité d'un courant d'injection) sur l'équipement jusqu'à provoquer une défaillance. Le test s'achève au niveau ayant provoqué la défaillance puisqu'on suppose que l'équipement restera susceptible aux niveaux supérieurs. Le seuil de l'équipement S_e ainsi testé se situe donc entre le dernier niveau de fonctionnement normal (état bas dans la Figure 4) i.e. $nt_e - \delta_e$ et le premier niveau de défaillance (état haut dans la Figure 4) i.e. nt_e. La variable δ_e est un scalaire correspondant à la différence entre les deux derniers seuils mesurés : celui ayant provoqué la défaillance et le précédent. La variable δ_e est indispensable puisqu'il n'est pas possible de tester en pratique tous les niveaux possibles. Cette approche expérimentale permet donc de récolter un échantillon de donnés regroupées dans un vecteur que l'on notera $\boldsymbol{Data}$. Ce vecteur contient *n* intervalles si $\delta_e > 0$ ou *n* seuils si $\delta_e = 0$. Dans le cas où $\delta_e > 0$, les intervalles récoltés sont notés $\Delta S_e = [nt_e - \delta_e, nt_e]$. Dans certains cas particuliers, comme en simulation, on peut observer directement le seuil i.e. $\delta_e = 0$. Dans ce cas, $\boldsymbol{Data}$ contient les seuils S_e. L'objectif final de cette approche expérimentale est de remonter à la distribution sous-jacente des seuils à partir de cet échantillon de données.

$$\boldsymbol{Data} = \begin{cases} [\Delta S_1, \dots \Delta S_e, \dots \Delta S_n] \ \forall \delta_e > 0 \\ [S_1, \dots S_e, \dots S_n] \ \forall \delta_e = 0 \end{cases} . \quad (1)$$

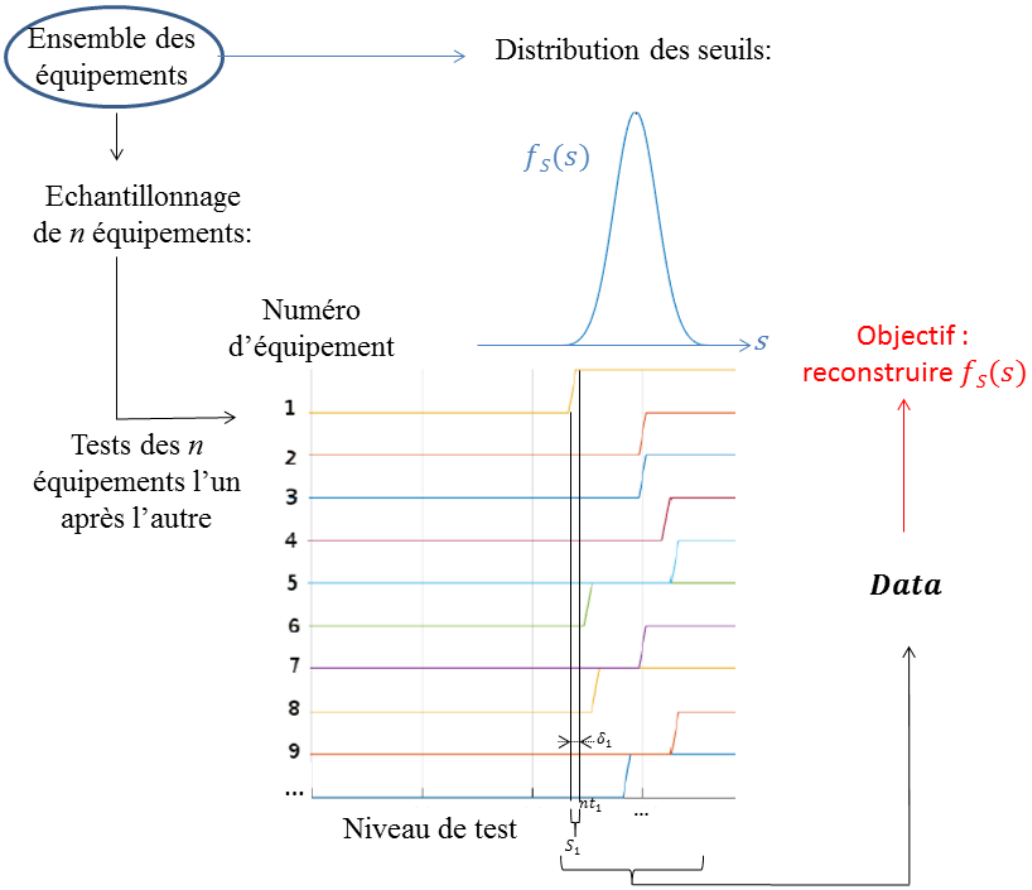

Figure 4 : Schéma de principe des tests de susceptibilité équipements. Etat haut = défaillance.

Le problème posé est en fait celui de l'inférence statistique : estimer des caractéristiques inconnues (nature et paramètres de la fonction de distribution sous-jacente) d'une population (ensemble des seuils de susceptibilité possibles) à partir d'un échantillon limité (mesures de susceptibilité) de la population. On peut distinguer deux types d'inférence paramétriques, l'Inférence Bayésienne (IB) et l'Inférence selon le Maximum de Vraisemblance (IMV). L'IB n'est pas encore couramment utilisée dans la communauté CEM. Récemment, l'IB a été envisagée pour traiter des incertitudes de mesures [27]. Yuhao et al. ont proposé d'utiliser une approche bayésienne pour estimer la distribution de susceptibilité à partir d'un ensemble limité d'expériences, en ne fournissant toutefois que la meilleure estimation (scalaire) des paramètres de la distribution [28].

2. Traitement des données mesurées par inférence statistique

a. Rappels sur les distributions et leurs paramètres

Une variable aléatoire, comme celle représentant le seuil de susceptibilité *S*, peut prendre plusieurs valeurs d'où le terme *variable* mais de manière non reproductible, d'où le terme *aléatoire*. Il est tout de même possible de quantifier la probabilité d'occurrence de chacune des valeurs possibles de *S* par des fonctions : la densité de probabilité ou la *probability distribution function* (*pdf*) et la fonction de répartition ou *cumulative distribution function* (*cdf*). Les abréviations *pdf* et *cdf* seront systématiquement utilisées par la suite. La *pdf* quantifie la probabilité d'observer une réalisation particulière notée *s* de la variable aléatoire *S*. Cette fonction est notée de façon générique $f_S(s)$. La *cdf* quant à elle, quantifie la probabilité d'observer une réalisation dans un intervalle donné. Elle est noté $F_S(s)$. Ces deux fonctions se déduisent l'une de l'autre puisque la *cdf* est l'intégrale de la *pdf* :

$$F_S(s) = \mathbb{P}(S \leq s) = \int_{-\infty}^{s} f_S(s)ds. \tag{2}$$

L'estimation empirique de la *cdf* étant plus facile, celle-ci est plus couramment utilisée que la *pdf*. La *cdf* et de la *pdf* ont une expression analytique connue pour les lois usuelles. Ces expressions sont paramétrées par un vecteur noté $\boldsymbol{\theta}$. Pour la plupart des distribution usuelles, comme toutes celles traitées ici, le vecteur $\boldsymbol{\theta}$ a deux composantes, θ_1 et θ_2. C'est notamment le cas lorsque la distribution est Normale (N), Log-Normale (LN) ou Weibull (W), lois qui seront utilisées dans la suite de ce chapitre. De plus, selon la loi, chaque composante représente une caractéristique particulière de la distribution observée. Par exemple lorsque la distribution est N, θ_1 est l'espérance μ et θ_2 est la variance σ^2. Pour une distribution W, θ_1 est appelé le paramètre d'échelle γ et θ_2 celui de la forme β. Lorsque S suit une loi N, l'expression analytique de la *cdf* de S est donnée en (3), tandis que si S suit une distribution W, sa *cdf* est donnée en (4).

$$F_S(s|\mu, \sigma^2) = \frac{1}{2}\left(1 + erf\left(\frac{s-\mu}{\sigma\sqrt{2}}\right)\right). \tag{3}$$

$$F_S(s|\gamma, \beta) = 1 - e^{-\left(\frac{s}{\beta}\right)^{\gamma}}. \tag{4}$$

L'objectif est de reconstruire la distribution de susceptibilité à partir de l'échantillon de mesure. Afin d'y parvenir, une méthode d'inférence peut être appliquée à l'échantillon. Il existe deux familles d'inférence : paramétrique (la plus classique) et non-paramétrique. Nous allons traiter uniquement la première et n'évoquer la deuxième que comme piste de réflexion en conclusion de ce chapitre (section VII). Une distribution est définie par :

- Une expression analytique de la *pdf*/*cdf* pour chaque loi supposée sous-jacente.
- Un vecteur paramètre $\boldsymbol{\theta}$ modulant la fonction *pdf*/*cdf*.

Lorsqu'une méthode paramétrique est utilisée, une expression de la fonction *pdf*/*cdf* analytique est supposée pour ensuite estimer le vecteur paramètre $\boldsymbol{\theta}$ la modulant. Il existe un grand nombre de distributions [29] et il n'est pas possible de choisir a priori la plus adaptée sans informations supplémentaires (expériences préalables, connaissances du phénomène physique à l'œuvre, symétrie de la distribution,…). Supposer une *pdf*/*cdf* est évidement un facteur limitant. Cependant, le paramétrage affecte beaucoup la forme de la fonction *pdf*/*cdf*, surtout pour certaines lois telle que la loi W comme le montre la Figure 5 et, dans une moindre mesure, pour la loi N comme indiqué par la Figure 6. De plus, toutes les lois sont plus ou moins fortement liées, ce qui transparait de la Figure 7 où certaines lois sont équivalentes pour des valeurs particulières de leurs paramètres. Cette figure, à la fois riche, synthétique et non-exhaustive est extraite de [8]. Toutefois, cette hypothèse a priori de distribution constitue inévitablement une première source d'erreur. La deuxième source d'erreur étant l'estimation du paramètre $\boldsymbol{\theta}$.

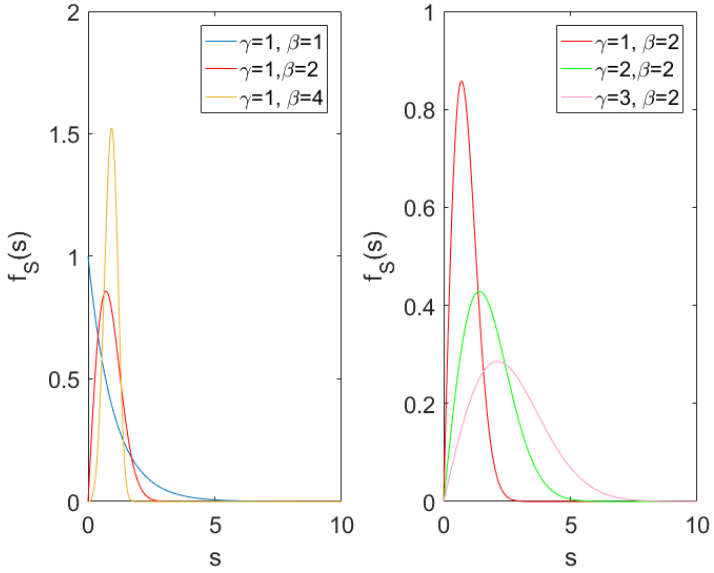

Figure 5: Effet des paramètres $\boldsymbol{\theta} = [\gamma, \beta]$ sur la pdf d'une loi de W.

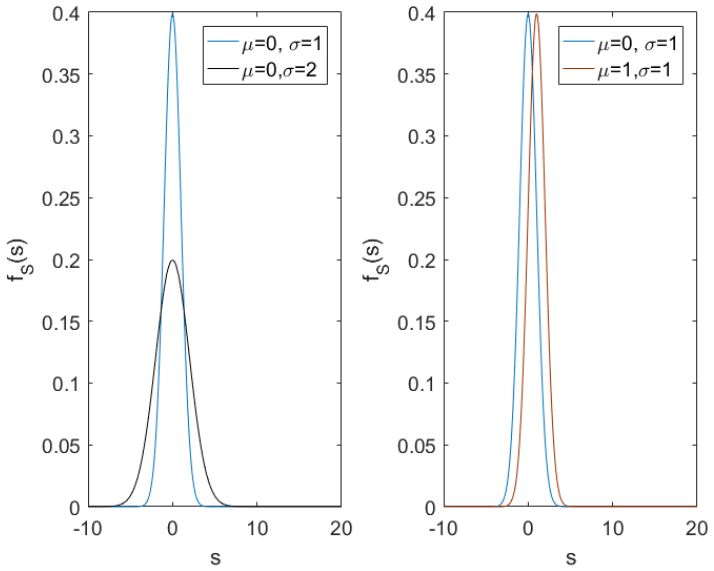

Figure 6 : Effet des paramètres $\boldsymbol{\theta} = [\mu, \sigma^2]$ sur la pdf d'une loi N.

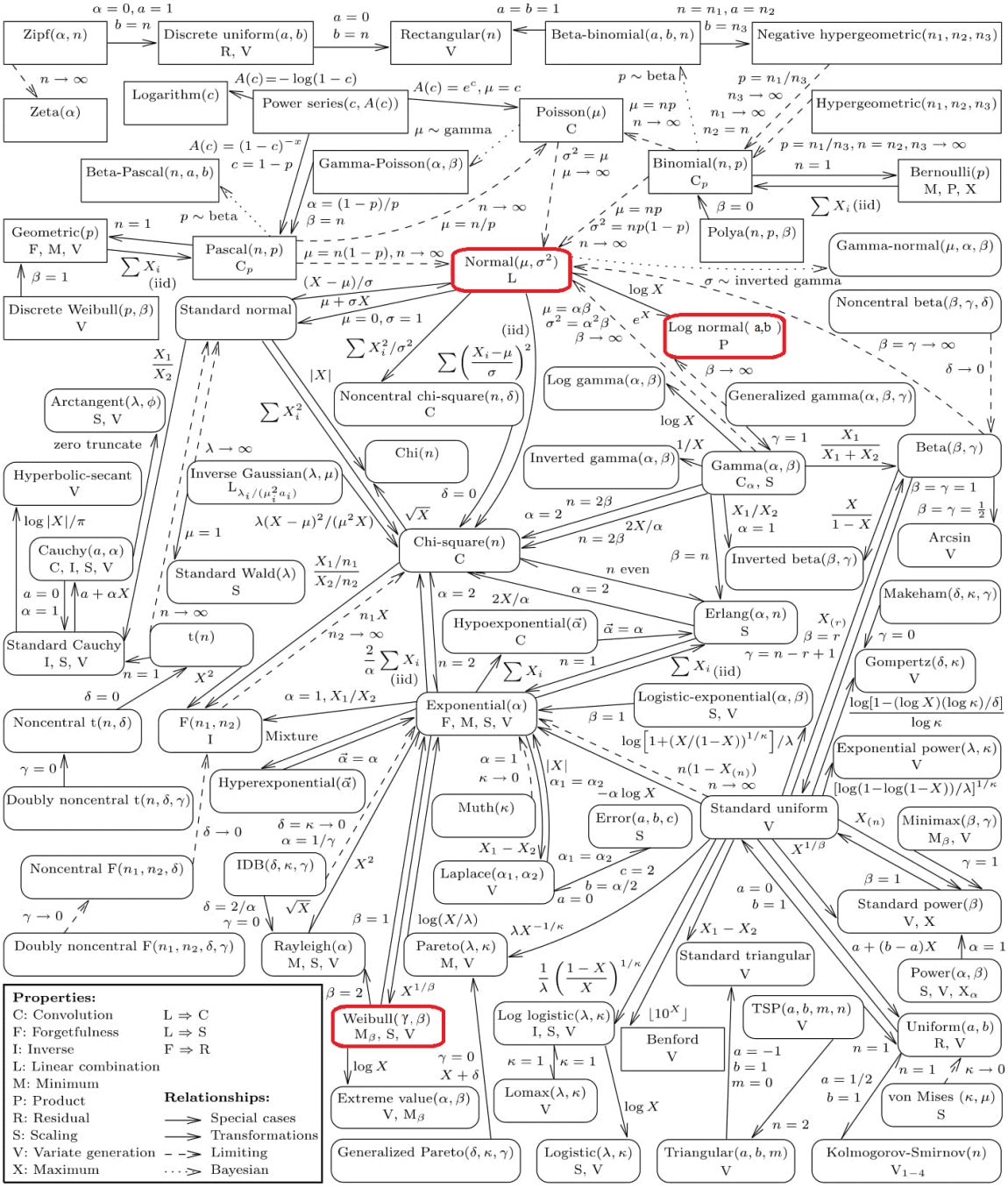

Figure 7 : Ensemble non exhaustif des distributions univariées existantes et leurs liens relatifs (soit par leurs paramètres, soit par transformation de variable). Figure extraite de [29]. Encadrés rouges = distributions utilisées dans ce chapitre.

b. Inférence paramétrique

Il existe deux principales méthodes d'inférence paramétrique, l'IMV et l'IB. Ces deux méthodes d'inférence diffèrent selon leur interprétation du vecteur de paramètres $\boldsymbol{\theta}$ de la fonction de la distribution supposée sous-jacente. En ce qui concerne l'IMV, le paramètre vectoriel $\boldsymbol{\theta}$ à estimer est déterministe alors que pour l'IB, il s'agit d'une variable aléatoire. Le résultat de l'IB est donc une distribution de probabilité de la variable aléatoire alors que le résultat de l'IMV est une fonction de vraisemblance. Toutefois, la fonction de vraisemblance est également exprimée en termes de probabilités et peut être interprétée comme une distribution de probabilité (son intégrale valant 1). Il

n'est pas nécessaire de spécifier pour l'IMV une distribution a priori de $\boldsymbol{\theta}$ puisque celle-ci est nécessairement uniforme. En revanche, avec l'IB, l'utilisateur est libre de choisir un a priori adéquat. Ainsi, l'IMV peut être considérée comme un cas particulier de l'IB.

A partir d'un échantillon formé de différentes mesures de seuils regroupées dans le vecteur $\boldsymbol{Data}$ défini dans (1), le théorème de Bayes permet d'estimer a posteriori la *pdf* du paramètre $\boldsymbol{\theta}$, appelée f_{post} :

$$f_{post}(\boldsymbol{\theta}|\boldsymbol{Data}) = \frac{f_{prior}(\boldsymbol{\theta}) f_{Like}(\boldsymbol{Data}|\boldsymbol{\theta})}{\int f_{prior}(\boldsymbol{\theta}) f_{Like}(\boldsymbol{Data}|\boldsymbol{\theta})\, d\boldsymbol{\theta}}, \tag{5}$$

où f_{prio} est la densité a priori de $\boldsymbol{\theta}$ tandis que f_{Like} est la vraisemblance d'avoir observé $\boldsymbol{Data}$ sachant $\boldsymbol{\theta}$. Le dénominateur en (5) est une opération de normalisation permettant d'obtenir une *pdf* (i.e. intégrale de f_{post} valant 1). Le maximum de f_{post} est un estimateur scalaire de $\boldsymbol{\theta}$ le plus probable.

Dans le cas de l'IMV, la densité a priori est nécessairement uniforme. L'opération de normalisation n'est plus nécessaire, l'intégrale de f_{post} valant déjà 1. Dans ce cas, la distribution a posteriori de $\boldsymbol{\theta}$ devient (à une constante près) :

$$f_{post}(\boldsymbol{\theta}|\boldsymbol{Data}) = f_{Like}(\boldsymbol{Data}|\boldsymbol{\theta}). \tag{6}$$

Une approche numérique calculant (5) pour l'IB ou (6) pour l'IMV est pertinente et repose sur les arguments suivants :

- Il est possible pour certaines distributions d'avoir une expression analytique de f_{post}, comme dans le cas N et LN. Lorsque cette expression existe et qu'elle est analytiquement dérivable, il est également possible d'obtenir analytiquement la densité a posteriori de l'estimateur $\boldsymbol{\theta}$ ainsi que sa valeur la plus probable. Ce n'est cependant pas toujours le cas, par exemple avec une distribution W, même avec un a priori uniforme.
- Enfin, calculer la valeur de normalisation, i.e. calculer l'intégrale figurant au dénominateur de (5), n'est pas non plus toujours réalisable.

Une approche numérique est donc beaucoup plus robuste. A notre connaissance, il n'existe pas de fonction Matlab calculant (5). Notre implémentation se trouve en Annexes.

i. Densité a priori

L'IB nécessite une distribution a priori des paramètres à estimer. Le choix de la distribution de l'a priori affecte de moins en moins la distribution a posteriori, à mesure que la taille de l'échantillon augmente. Il est possible de tirer profit des connaissances tirées d'expériences antérieures. Toutefois, si aucune connaissance n'est disponible, il faut choisir un a priori non informatif. Pour trouver de tels a priori, deux conditions doivent être remplies. La première est l'entropie maximale et la seconde est l'invariance de l'échelle spatiale des paramètres [30][31]. Le choix de l'a priori dépend de la distribution supposée sous-jacente et de la méthode utilisée pour le trouver. Il existe trois méthodes pour obtenir des a priori non informatifs :

- Le *Maximal Data Information Prior* (MDIP) de Zellner qui est basé sur la maximisation d'un critère d'information [32].
- L'a priori de Jeffrey qui est basé sur la matrice d'information de Fisher [33].
- L'a priori dit de référence qui est une modification de l'a priori de Jeffrey [34].

Pour une distribution sous-jacente donnée, plus d'une distribution a priori non-informative est possible et plusieurs méthodes peuvent conduire au même a priori. Un catalogue d'a priori existe pour les distributions usuelles [35]. Les densités a priori que nous avons utilisées se trouvent dans le

Tableau 5. Nous avons utilisé le MDIP pour la distribution N (qui est le même que l'a priori dit de référence). L'a priori de Jeffrey est utilisé pour la distribution W (qui est aussi la même que l'a priori de référence). Pour la distribution LN, dans [36] la densité a priori est une generalized inverse Gaussian. Nous avons choisi l'a priori de Jeffrey comme dans [37].

Tableau 5 : Distributions a priori utilisées selon la loi sous-jacente supposée.

Loi supposée	N	LN	W
Densité a priori	$f_{prior}(\mu,\sigma^2)=\frac{1}{\sigma^2}$	$f_{prior}(a,b^2)=\frac{1}{b^2}$	$f_{prior}(\gamma,\beta)=\frac{1}{\gamma\beta}$

ii. <u>Fonction de vraisemblance</u>

La fonction de vraisemblance s'écrit :

$$f_{Like}(\boldsymbol{Data}|\boldsymbol{\theta})=\prod_{e=1}^{n}F_S(nt_e)-F_S(nt_e-\delta_e), \tag{7}$$

où F_S est la *cdf* de la distribution hypothétique et nt_e le premier niveau de contrainte appliqué provoquant une défaillance du e^{eme} équipement alors qu'il fonctionnait normalement au niveau précédemment testé, i.e. $nt_e-\delta_e$. Dans certaines situations, comme en simulation $\delta_e=0\ \forall e$, le seuil mesuré est exactement la réalisation aléatoire de *S*. Dans ce cas $\boldsymbol{Data}=[S_1,\dots,S_n]$ et la fonction de vraisemblance devient :

$$f_{Like}(\boldsymbol{Data}|\boldsymbol{\theta})=\prod_{e=1}^{n}f_S(S_e). \tag{8}$$

Pour quelques distributions usuelles, il est possible d'obtenir une expression analytique de la fonction de vraisemblance. De plus, pour certaines distributions, il est possible d'accéder à une solution analytique qui maximise la fonction de vraisemblance en calculant sa dérivée. C'est le cas pour les distributions gaussienne (N) et Log-normale (LN) [38][39]. Dans le casd'une distribution de Weibull (W), une approche numérique et nécessaire, mais des approximations sont disponibles [40][41]. Ces estimateurs sont rassemblés dans les Tableau 6 pour la distribution N et Tableau 7 pour la distribution de W.

Tableau 6 : Estimateurs les plus probables lorsque la distribution est N avec une distribution a priori de $\boldsymbol{\theta} = [\mu, \sigma^2]$ uniforme (i.e. estimateur scalaire de l'IMV). Deux estimateurs sont possibles pour θ_2 : le premier ($\widehat{\sigma_n^2}$) est biaisé alors que le deuxième ($\widehat{\sigma_{n-1}^2}$) ne l'est pas.

Estimateur	**Expression**	
$\hat{\theta}_1$	$\hat{\mu} = \frac{1}{n}\sum_{e=1}^{n} S_e$	
$E[\hat{\theta}_1]$	$E[\hat{\mu}] = \mu$	
$V[\hat{\theta}_1]$	$V[\hat{\mu}] = \frac{\sigma^2}{n}$	
$\hat{\theta}_2$	$\widehat{\sigma_n^2} = \frac{1}{n}\sum_{e=1}^{n}(S_e - \hat{\mu})^2$	$\widehat{\sigma_{n-1}^2} = \frac{1}{n-1}\sum_{e=1}^{n}(S_e - \hat{\mu})^2$
$E[\hat{\theta}_2]$	$E\left[\widehat{\sigma_n^2}\right] = \frac{n-1}{n}\sigma^2$	$E\left[\widehat{\sigma_{n-1}^2}\right] = \sigma^2$
$V[\hat{\theta}_2]$	$V[\widehat{\sigma_n^2}] = \frac{2(n-1)}{n^2}\sigma^4$	$V[\widehat{\sigma_{n-1}^2}] = \frac{2}{(n-1)^2}\sigma^4$

On remarquera que :

$$V[\sigma_{n-1}^2] = \left(\frac{n}{n-1}\right)^2 V[\sigma_n^2]. \tag{9}$$

Lorsque la distribution est LN, les estimateurs sont les mêmes que dans le cas N après avoir appliqué le logarithme sur chaque réalisation de l'échantillon [38]. Ainsi, l'estimateur de θ_1 est :

$$\hat{a} = \frac{1}{n}\sum_{e=1}^{e} ln(S_e), \tag{10}$$

tandis que celui de θ_2 est :

$$\widehat{b^2} = \frac{1}{n}\sum_{i=1}^{n}(ln(S_e) - \hat{a})^2. \tag{11}$$

Tableau 7 : Estimateurs les plus probables lorsque la distribution est W avec une distribution a priori de $\boldsymbol{\theta} = [\gamma, \beta]$ uniforme (i.e. estimateur scalaire de l'IMV).

Estimateur	**Expression**
$\hat{\theta}_1$	$\hat{\gamma} = \left(\frac{1}{n}\sum_{e=1}^{n} S_e^{\hat{\beta}}\right)^{1/\hat{\beta}}$
$E[\hat{\theta}_1]$	$E[\hat{\gamma}] \neq \gamma$
$V[\hat{\theta}_1]$	$V[\hat{\gamma}] = 1.108665\frac{\gamma^2}{n\beta^2}$
$\hat{\theta}_2$	$\hat{\beta} = \left(\left(\sum_{e=1}^{n} S_e^{\hat{\beta}} \ln s_i\right)\left(\sum_{e=1}^{n} S_e^{\hat{\beta}}\right)^{-1} - \frac{1}{n}\sum_{e=1}^{n} \ln \mathrm{S_e}\right)^{-1}$
$E[\hat{\theta}_2]$	$E[\hat{\beta}] \neq \beta$
$V[\hat{\theta}_2]$	$V[\hat{\beta}] = 0.607927\frac{\beta^2}{n}$

Pour déterminer les estimateurs γ et β, une résolution numérique est nécessaire. Des approximations sont fournies dans le Tableau 7, pouvant servir de valeur de départ d'algorithmes de recherche itérative (fonction Matlab *vpasolve)*. On peut aussi tirer aléatoirement (loi uniforme) et tester plusieurs valeurs de départ. De plus, les estimateurs sont biaisés. La correction du biais nécessite une approche empirique par tables de Monte Carlo [42]. Nous choisissons de conserver l'estimateur non corrigé car ces tables ne sont pas connues de Matlab. L'utilisation des estimateurs non-biaisés aurait nécessité une longue étude supplémentaire et indépendante, dédiée à la génération automatiquement de ces tables. De plus, les estimateurs $\hat{\gamma}$ et $\hat{\beta}$ ne sont pas indépendants :

$$cov(\hat{\gamma}, \hat{\beta}) = 0.257022\frac{\gamma}{\beta}. \tag{12}$$

iii. <u>Encadrement probabiliste de la probabilité de défaillance</u>

Les deux méthodes d'inférence fournissent une distribution f_{post} du paramètre estimé. La probabilité de défaillance $p_f(s)$ est la probabilité d'avoir une défaillance si le niveau de test atteint la valeur s. Une fois que les paramètres $\boldsymbol{\theta}$ sont estimés, $p_f(s)$ se calcule ainsi :

$$p_f(s) = F_S(s|\boldsymbol{\theta}). \tag{13}$$

Cependant, $\boldsymbol{\theta}$ est distribué selon la densité multivariée f_{post}. Chaque couple de paramètres (θ_1, θ_2) a une probabilité proportionnelle à f_{post} évaluée en θ_1, θ_2. Ainsi, à un niveau s, $p_f(s)$ suit lui-même une distribution. Par conséquent, $p_f(s)$ est associé à un faisceau de *cdf* probables. Il est possible de délimiter ce faisceau entre une borne supérieure $U(s)$ et une borne inférieure $L(s)$ avec un intervalle de confiance (ici choisi arbitrairement à 95 %). Les bornes sont donc les quantiles de $p_f(s)$. Par conséquent:

$$\mathbb{P}\big(p_f(s) \in [L(s), U(s)]\big) = 95\ \%. \tag{14}$$

Au-delà de l'examen de la probabilité de défaillance $p_f(s)$, les ingénieurs CEM peuvent être plus intéressés par une quantité qui en découle : le quantile s_α de susceptibilité associé à la probabilité

α. Cette quantité correspond au niveau de test auquel une probabilité de défaillance de α est atteinte, i.e. :

$$s_\alpha = F_S^{-1}(\alpha|\boldsymbol{\theta}). \tag{15}$$

Puisque $\boldsymbol{\theta}$ est une distribution, s_α suit aussi une distribution. C'est pourquoi nous introduisons un intervalle de confiance $[s_{low}, s_{upp}]$ encadrant s_α:

$$\begin{cases} s_{low} = U^{-1}(\alpha) \\ s_{upp} = L^{-1}(\alpha) \end{cases}, \tag{16}$$

avec une confiance de 95 %. Ainsi :

$$\mathbb{P}\{s_\alpha \in [s_{low}, s_{upp}\} = 95\ \% \tag{17}$$

Dans le contexte de l'estimation de la probabilité de défaillance d'un équipement agressé par une IEMI, une estimation correcte de la queue inférieure de la distribution de susceptibilité est particulièrement importante. En effet, dans un tel scénario, la distribution estimée de la contrainte générée par l'IEMI est analysée en lien avec la distribution estimée de la susceptibilité. Dans de nombreux cas pratiques le risque de défaillance de l'équipement agressé est principalement déterminé par la queue supérieure de la distribution de la contrainte et la queue inférieure de la distribution de la susceptibilité. C'est pourquoi nous examinerons la qualité de l'estimation du seuil de susceptibilité dans la queue de distribution, avec $\alpha = 5\ \%$.

III. Estimation du type de distribution

Dans cette section, nous voulons déterminer quelle inférence (IB ou IMV) pour de petits échantillons, convient le mieux pour choisir la nature de la fonction de distribution parmi plusieurs distributions possibles. Dans le contexte de la CEM, nous réduisons les distributions de susceptibilité possibles aux fonctions de distribution N, LN et de W. Cette sélection repose sur les arguments suivants :

- La fonction de distribution N, paramétrée par (μ, σ^2), représente un processus résultant de contributions additives de multiples variables aléatoires indépendantes.
- La fonction LN, paramétrée par (a, b^2), liée à la loi N (voir la Figure 7), peut être trouvée dans les problèmes CEM impliquant la diaphonie. L'amplitude de la tension sur le fil victime (ou piste) est proportionnelle à l'inverse du logarithme de la distance entre les deux fils concernés.
- La fonction W, paramétrée par (γ, β), est très souvent rencontrée dans l'évaluation des probabilités de défaillance. C'est aussi une fonction de distribution générale qui englobe d'autres fonctions telles que les distributions Exponentielle et de Rayleigh.

Cette section vise à comparer les performances d'un test statistique d'ajustement d'un critère de sélection BIC (Bayesian Inference Criterion) [43] pour choisir parmi ces trois possibilités.

Tout d'abord, nous étudions la proximité relative entre ces trois distributions (sous-section 1), puis nous rappelons le principe des tests statistiques (sous-section 2) ainsi que celui du BIC (sous-section 3). Enfin les résultats de la comparaison sont présentés et analysés (sous-section 4).

1. Coefficient de variation et proximité entre distributions

Dans l'analyse qui va suivre, la moyenne μ de l'échantillon est fixe et arbitrairement choisie égale à 10. Seul l'écart-type est ajusté afin d'obtenir une variation du coefficient de variation (CV) de 1 à 100 %, sachant que $CV = \sigma/\mu$. Les trois distributions sont paramétrées pour atteindre le CV prescrit. Ensuite, nous calculons la distance moyenne entre leurs *cdf* (exprimée en valeur absolue). Nous

traçons sur la Figure 8 les distances entre W et N (noté W,N), LN et W (LN,W) et enfin entre N et LN (N,LN). Nous remarquons que les distances W,N et LN,W augmentent avec le CV. C'est également vrai pour W,N qui passe cependant par un minimum local pour un CV de l'ordre de 32.6 %.

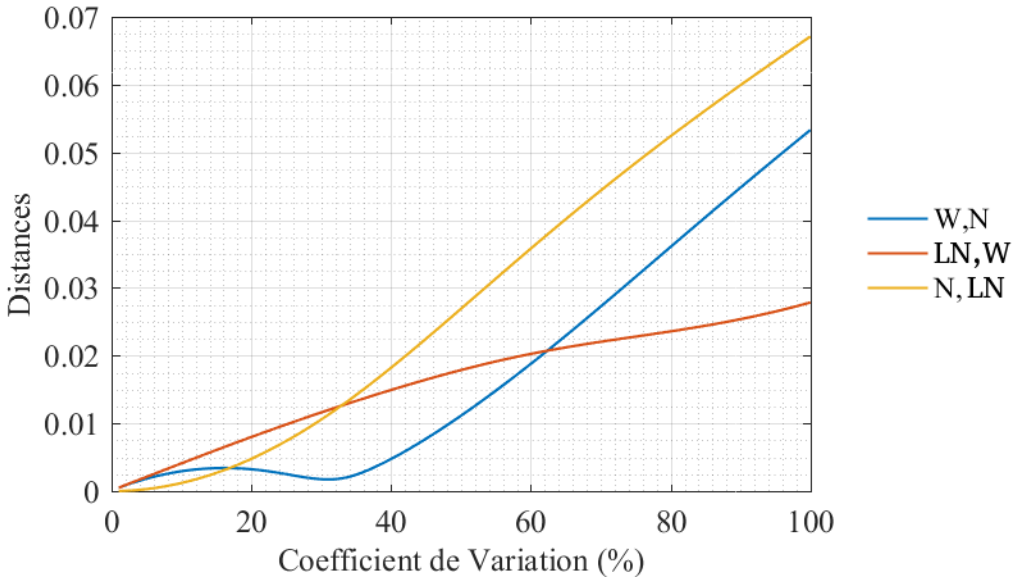

Figure 8 : Distance moyenne (en absolue) entre les trois distributions.

Le CV influencera donc la qualité de l'estimation de la distribution sous-jacente. En particulier avec un CV autour de 32.6 %, la distance entre N et W est très faible. Il n'est plus pertinent d'essayer de les distinguer car N et W sont similaires dans ce cas. Nous avons choisi un CV=62 %, une valeur réaliste dans le contexte de la susceptibilité d'équipements électroniques.

2. Tests statistiques

Les tests statistiques regroupent plusieurs familles selon l'objectif du test (ajustement de loi, indépendance, …). Nous nous intéressons ici à la famille des tests paramétriques d'ajustement de loi [44].

Un test statistique quantifie l'accord entre un modèle statistique et les données disponibles. La procédure d'un test est en générale la suivante. On commence par formuler l'hypothèse nulle, notée H_0. On décide aussi d'un facteur de prise de risque ζ. L'hypothèse concerne le type de loi, i.e. on suppose que les échantillons proviennent d'un tirage aléatoire suivant la loi postulée par H_0. On note son contraire H_1, c'est-à-dire l'hypothèse qui consiste à rejeter H_0. Par exemple, H_0 =*La loi est N* et H_1= *la loi n'est pas N*.

A partir de l'échantillon, on calcule la statistique du test qui dépend du test choisi. Cette statistique est une variable aléatoire car elle est calculée à partir d'un échantillon, issu lui-même d'un tirage aléatoire. Cependant, la statistique suit généralement une loi connue. Ainsi, une étude analytique ou par Monte-Carlo fournit les quantiles de la statistique lorsque l'on sait que H_0 est vraie. Ceux-ci sont en général donnés dans des tables selon la taille de l'échantillon. Si la probabilité d'observer la statistique calculée est inférieure à ζ alors le test rejette H_0 (on choisit H_1). Il est donc possible de choisir H_1 alors que H_0 est vrai avec une probabilité de ζ (erreur de type I notée E_I). Si la probabilité d'observer cette statistique est supérieure à ζ le test ne peut pas rejeter H_0 (on choisit H_0). Il est donc possible de choisir H_0 alors que H_1 est vraie (erreur de type II notée E_{II}). La quantité 1- E_{II} est appelée la puissance du test. Lorsque la puissance du test est faible on dit que le test est conservateur.

Du fait de leur construction, il est facile d'évaluer l'erreur de type I qui tend vers ζ lorsque la taille de l'échantillon est suffisamment grande. Cependant il n'est pas possible d'estimer a priori

l'erreur de type II. Mathématiquement, les deux erreurs s'écrivent :

$$E_I = \mathbb{P}(H_1|H_0) \underset{n}{\rightarrow} \xi, \tag{18}$$

$$E_{II} = \mathbb{P}(H_0|H_1). \tag{19}$$

Il existe de très nombreux tests comme : Kolmogorov-Smirnov [45], Anderson-Darling (AD) [46], Lilliefors [47] et Shapiro-Wilk [48][49]. Une étude a comparé les performances de ces différents tests [50]. Pour de grands échantillons, Shapiro-Wilk, est le plus performant en termes de puissance. Cependant, pour de petites tailles échantillons (quelques dizaines), tous les tests comparés sont équivalents mais très conservateurs. De plus, les tables existantes pour analyser un test de Shapiro-Wilk, ne sont valables que pour une loi N. Plus récemment encore (2018), une nouvelle étude [51] a comparé 18 tests avec pour H_0 l'appartenance à la loi N. Il ressort de cette étude que les différences de performances entre les tests ne sont pas clairement établies pour les petites tailles à l'exception notable du test de Kolmogorov-Smirnov qui est clairement le moins performant en terme d'E_{II}. Nous n'avons donc pas choisi Kolmogorov-Smirnov. De plus, nous étudierons d'autres distributions que N donc nous ne choisissons pas Shapiro-Wilk. Finalement, notre choix se porte sur le test d'Anderson Darling (AD).

Le test d'AD se base la distance entre une *cdf* postulée *F* (dont les paramètres sont estimés par l'IMV), la *cdf* empirique et des poids de pondération afin de donner plus d'importances aux observations extrêmes. La statistique d'AD est la suivante :

$$A_n^2 = -n - \sum_{e=1}^{n} \frac{2e-1}{n}\Big(ln\big(F_S(S_e)\big) + ln\big(1 - F_S(S_{n+1-e})\big)\Big), \tag{20}$$

où $S_1, \dots, S_e, \dots, S_n$ sont les seuils de susceptibilité ordonnés (i.e. $S_e \leq S_{e+1}$). La fonction Matlab *adtest* permet d'utiliser le test d'AD.

3. Test du BIC

Le BIC quantifie la vraisemblance (en tant que probabilité) de l'estimation de la *cdf* (IB ou IMV) en fonction des données (*cdf* empirique). En général, le BIC est utilisé pour discriminer différents modèles ayant pour nombre total de paramètres k' en fonction d'un ensemble de données de taille n. Il existe d'autres critères à cette fin comme l'AIC [43]. La taille de l'échantillon n et le nombre de paramètres k' interviennent dans le calcul de ces critères pour pénaliser les modèles plus complexes et choisir le plus simple (plus petit nombre de paramètres k'). Dans notre cas, les trois distributions ont le même nombre de paramètres (k'=2). Par conséquent, le BIC et l'AIC sont équivalents. Nous choisissons arbitrairement le BIC. Le BIC est calculé comme suit :

$$BIC = ln(n)\, k' - 2ln(l), \tag{21}$$

où l est la fonction de vraisemblance exprimée en fonction de la réalisation (S_e, e=1,...,n) de la variable aléatoire S (seuils de susceptibilité) et de l'estimation la plus probable. Comme les réalisations sont indépendantes :

$$l = \prod_{e=1}^{n} \mathbb{P}(S = S_e|\boldsymbol{\theta}) = \prod_{e=1}^{n} f_{Bino}\big(k_e|n, p_f(S_e)\big), \tag{22}$$

où f_{Bino} est la fonction de probabilité binomiale. Ses paramètres sont le nombre d'équipements défaillants k_e au niveau de contrainte S_e, n la taille de l'échantillon et $p_f(Se)$, la probabilité estimée de défaillance à partir de l'inférence. Lorsque S=S_e le nombre k_e de l'équipement est égal à la *cdf* empirique de S évaluée en S_e.

L'hypothèse conduisant au plus petit BIC est donc l'hypothèse la plus probable et sera donc celle retenue.

4. Comparaison AD/BIC

Afin de choisir une hypothèse de loi sous-jacente, trois tests sont comparés : AD, BIC se basant sur l'estimateur scalaire de $\boldsymbol{\theta}$ d'après l'IB, et BIC d'après l'estimateur scalaire de $\boldsymbol{\theta}$ fourni par l'IMV. Le critère de performance choisi est la probabilité d'un test à choisir la bonne hypothèse (e.g. choisir H_N si la loi est effectivement N). La procédure de comparaison est la suivante :

1. Tirage aléatoire d'un échantillon de taille n à partir de la distribution réelle connue a priori.
2. Pour chacune des trois hypothèses nulles (i.e. H_N, H_{LN} et H_W) :
    a. Appliquer une méthode d'inférence afin de calculer un estimateur scalaire de $\boldsymbol{\theta}$. Par exemple sous H_N :
        - Avec l'IMV, utiliser les estimateurs du Tableau 6.
        - Avec l'IB, calculer (5) avec l'a priori du Tableau 5 puis localiser le sommet de la distribution a posteriori de $\boldsymbol{\theta}$.
    b. Appliquer le BIC(IMV), BIC(IB) et l'AD avec l'estimateur calculé précédemment. Chaque test sélectionne une hypothèse de distribution.
3. Répéter 1000 fois les étapes 1 et 2 pour estimer la probabilité de choisir la bonne distribution.

Les résultats sont présentés pour plusieurs tailles d'échantillons n au sein de la Figure 9 lorsque la distribution est N et en Figure 10 lorsque la distribution est LN (jaune) et W (vert).

Quel que soit le test, plusieurs remarques peuvent être faites. Les proportions augmentent avec n et tendent comme attendu vers 100 %. Les proportions convergent plus rapidement vers 100 % dans la Figure 9 que dans la Figure 10. En effet, si au moins une réalisation négative est présente dans l'échantillon, ce qui n'est possible qu'avec une loi N, l'hypothèse sélectionnée est nécessairement H_N. Les résultats de l'AD et du BIC (IMV) sont relativement proches puisqu'ils utilisent tous deux l'IMV. De plus, si n est assez grand (≈ 40 dans Figure 9 et ≈ 80 dans Figure 10) tous les tests sont identiques.

Lorsque la distribution est N (Figure 9) et LN (en jaune dans Figure 10), la probabilité de choisir H_N la plus élevée provient d'AD, suivi de près par le BIC(IMV) alors que celle du BIC(IB) pour de très petits échantillons est très faible. Pour les distributions N et LN, l'IMV est donc plus performante que l'IB.

Lorsque la distribution est W (en vert dans Figure 10), le BIC(IB) aboutit à la probabilité la plus élevée, même pour les très petits n, alors que l'AD et le BIC(IMV) échouent la plupart du temps. Il est intéressant de noter que la probabilité de choisir H_W avec le BIC(IB) décroit localement lorsque n augmente (pour n <30). Pour la distribution W, il est préférable d'utiliser l'IB.

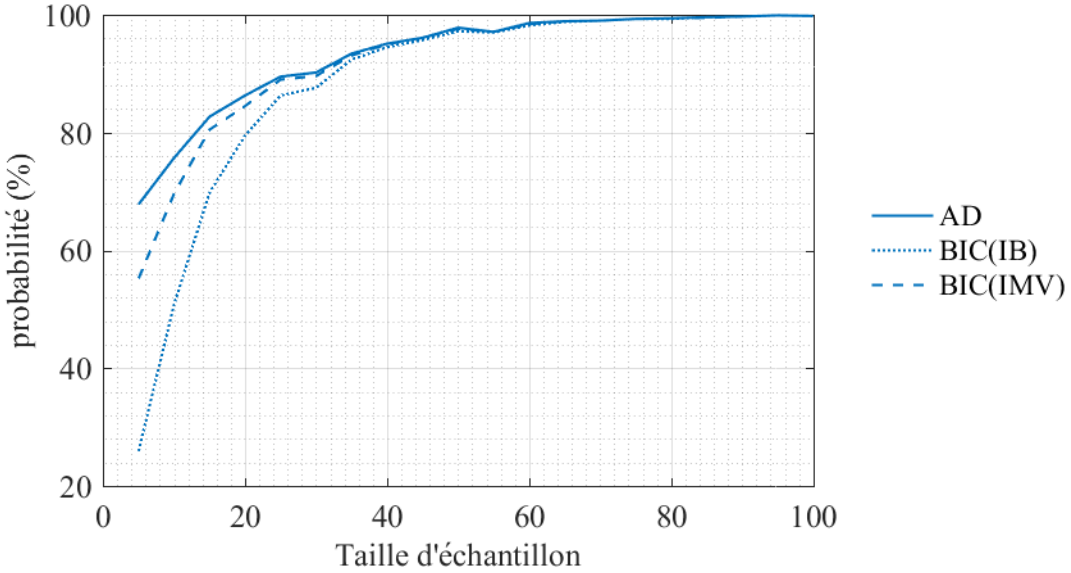

Figure 9 : Probabilité de choisir la loi N si la distribution vraie est effectivement N.

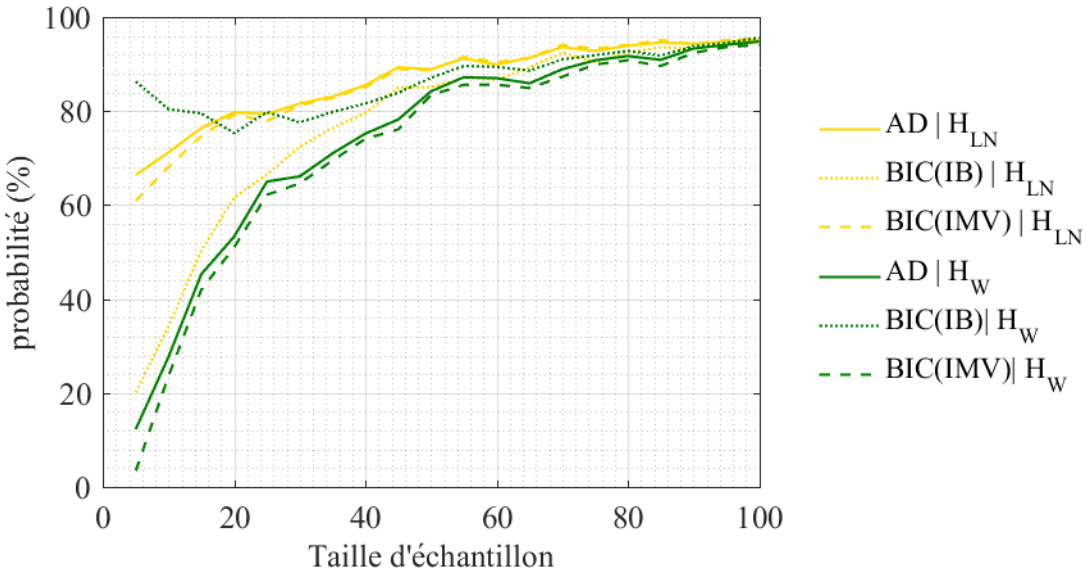

Figure 10 : Probabilité de choisir la loi LN (resp. W) alors que la loi vraie est LN (resp. W).

IV. Estimation des paramètres de la distribution

Dans cette section, nous comparons les deux méthodes d'inférence lorsque la distribution réelle est de type N ou W. Les paramètres réels de la distribution sous-jacente sont fixés de telle sorte que le CV soit égal à 20 %. Dans cette section, nous supposons que les deux inférences utilisent la bonne hypothèse (N dans la sous-section 1 et W dans la sous-section 2). Nous comparons d'abord l'estimation des deux composantes scalaires de $\boldsymbol{\theta}$. Puis, nous comparons les bornes encadrant s_α. La procédure utilisée pour comparer les estimateurs est la suivante :

1. Tirage aléatoire d'un échantillon de taille *n* à partir de la distribution réelle.
2. Sachant la nature de la distribution (i.e. N, LN ou W), application d'une méthode d'inférence afin de calculer un estimateur scalaire de $\boldsymbol{\theta}$. Par exemple, dans le cas N :
   - Avec l'IMV, utiliser les estimateurs du Tableau 6.
   - Avec l'IB, calculer (5) avec l'a priori figurant dans le Tableau 5 puis localiser le sommet

de la distribution a posteriori de $\boldsymbol{\theta}$.

En répétant 1000 fois cette procédure pour chaque taille *n*, nous pouvons très correctement estimer la moyenne et l'écart type de ces estimateurs de $\theta_1$et θ_2.

1. Cas de la distribution Normale

Dans la Figure 11, la moyenne et l'écart type de l'estimateur de θ_1 sont tracés pour les deux méthodes d'inférence. Les estimateurs (par l'IB et l'IMV) de θ_1 sont exactement les mêmes à la fois en moyenne et en écart type. En effet, la distribution marginale a priori de θ_1 est uniforme. Par conséquent, le choix de l'inférence n'a pas d'importance pour estimer θ_1.

Dans la Figure 12, la moyenne et l'écart type de l'estimateur de θ_2 sont tracés. La moyenne avec l'IB est systématiquement plus petite que la moyenne obtenue de l'IMV. Les deux estimations sont cependant inférieures à la valeur réelle. L'écart-type avec l'IB est inférieur à celui de l'IMV. L'estimation est donc un peu plus précise avec l'IB, surtout pour les petits *n*. Ainsi, l'IB entraîne un biais plus important mais une variance plus faible.

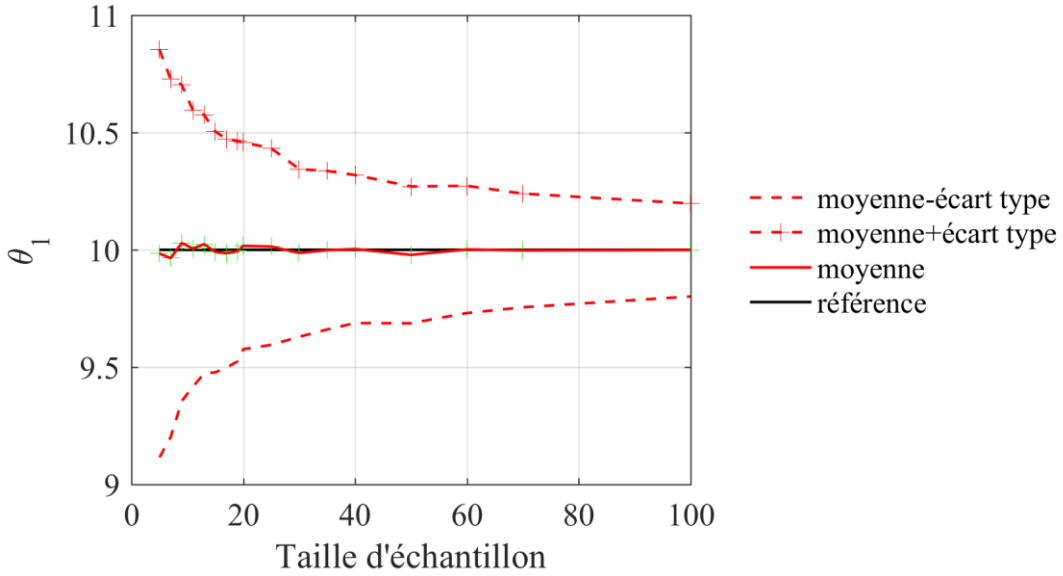

Figure 11 : Loi Normale, estimation du paramètre θ_1.

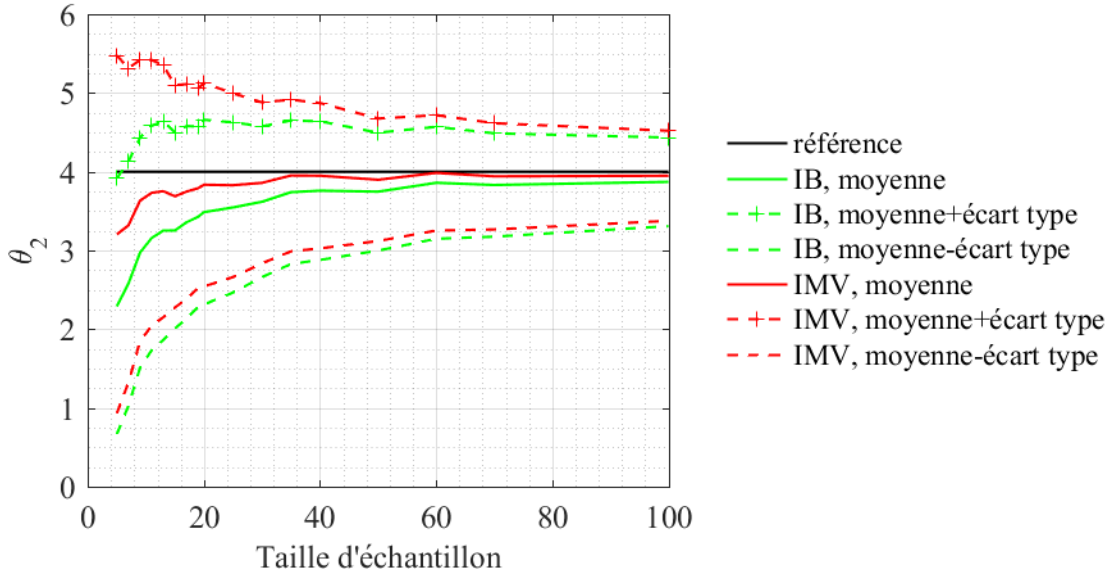

Figure 12 : Loi Normale, estimation du paramètre θ_2.

2. Cas de la distribution de Weibull

Dans la Figure 13 nous traçons la moyenne et l'écart type de l'estimateur de θ_1 pour les deux méthodes d'inférence. La moyenne de l'estimateur de θ_1 de l'IMV est légèrement plus proche de la valeur réelle que celle fournie par l'IB. Les écarts-types sont similaires. Par conséquent, l'IMV est plus pertinente que l'IB pour estimer θ_1.

Dans la Figure 14, nous traçons la moyenne et l'écart type de l'estimateur de θ_2. Dans ce cas, la moyenne de l'estimation de θ_2 avec l'IB est plus proche de la valeur vraie qu'avec l'IMV au prix d'une augmentation de l'écart-type. Tout comme le cas N, le choix entre l'IB et l'IMV pour l'estimation de θ_2 est un compromis entre biais et variance.

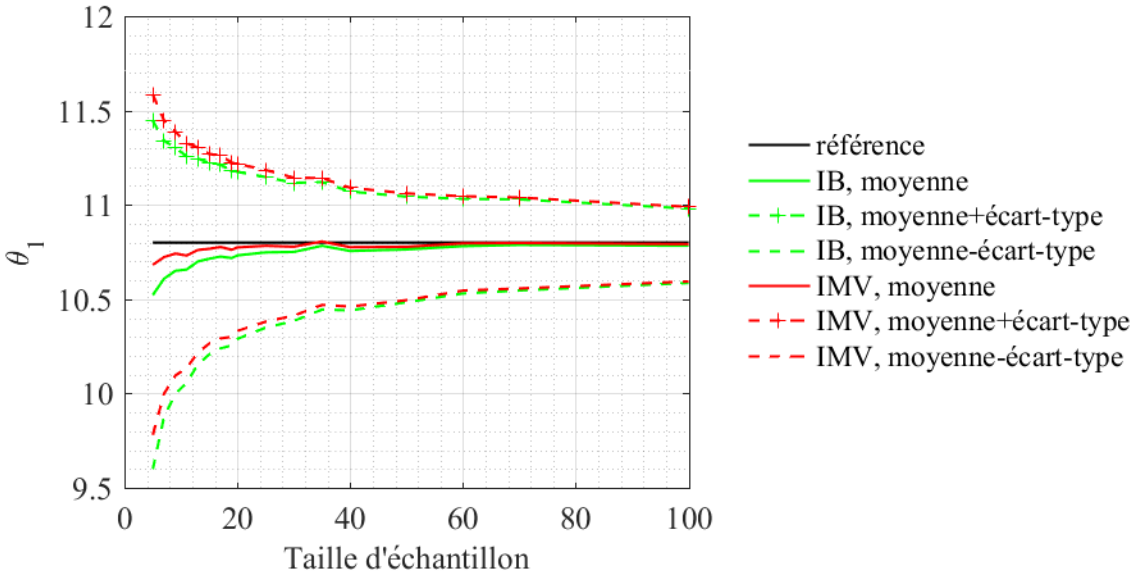

Figure 13 : Loi W, estimation du paramètre θ_1.

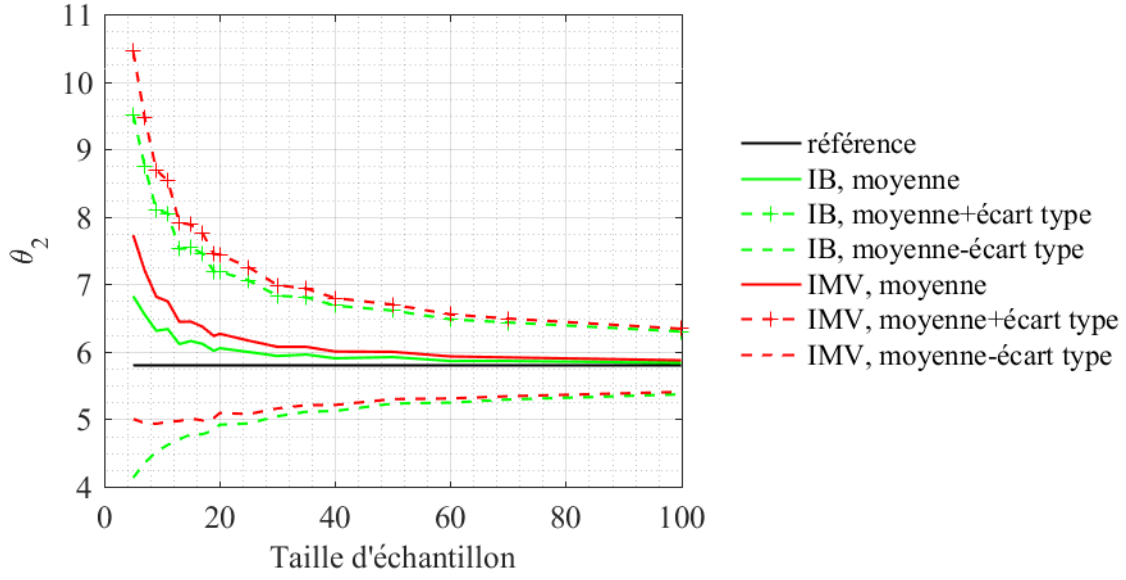

Figure 14 : Loi W, estimation du paramètre θ_2.

3. Conclusion sur l'estimation des paramètres

Comme attendu, les estimateurs des deux méthodes d'inférence convergent de manière similaire vers la valeur vraie des paramètres pour une taille d'échantillon supérieure à 40 environ. L'intérêt de l'IB est donc cantonné comme attendu aux petites tailles d'échantillon. Son intérêt essentiel est de limiter la variance de l'estimation des paramètres en contrepartie d'une augmentation du biais d'estimation dans le cas N.

V. Encadrement probabiliste du seuil de susceptibilité

Dans les deux sections précédentes (III et IV), nous avons utilisé l'estimateur scalaire (le plus probable) de $\boldsymbol{\theta}$. A l'heure actuelle, la majorité des études traitant de la susceptibilité en CEM s'arrêtent là, n'exploitant pas tout le potentiel des méthodes d'inférences. En effet, comme nous l'avons rappelé (section II. Etat de l'art), il est possible de prendre en compte l'ensemble des valeurs possibles de $\boldsymbol{\theta}$ (pas seulement les plus probable) afin d'encadrer la *cdf* de la susceptibilité.

Cette section présente une vérification de l'encadrement (17) (sous-section 1) puis une comparaison de l'estimation de s_{low} et s_{upp} par l'IB et l'IMV (sous-sections 2 et 3).

La procédure suivie est la suivante :

1. Générer un échantillon à partir de la distribution réelle.
2. Calculer (5) avec l'a priori selon la distribution supposée fournie dans le Tableau 5. Pour l'IMV l'a priori est uniforme.
3. Calculer les valeurs possibles de p_f selon (13).
4. A partir de la distribution de p_f en déduire un intervalle s_{low} et s_{upp} contenant le s_α réel à un degré de confiance (95 %) avec (16) et (14).

1. Validation de l'encadrement

Comme explicité précédemment (section II, sous-section 2), il est possible d'encadrer $p_f(s)$ par des intervalles de confiance. Ces bornes, U et L, dépendent de l'échantillon. Pour illustrer ce phénomène, nous avons généré 5 échantillons de petites tailles (n=5) à partir de la distribution N et avons tracé à la Figure 15, les 5 bornes supérieures et inférieures.

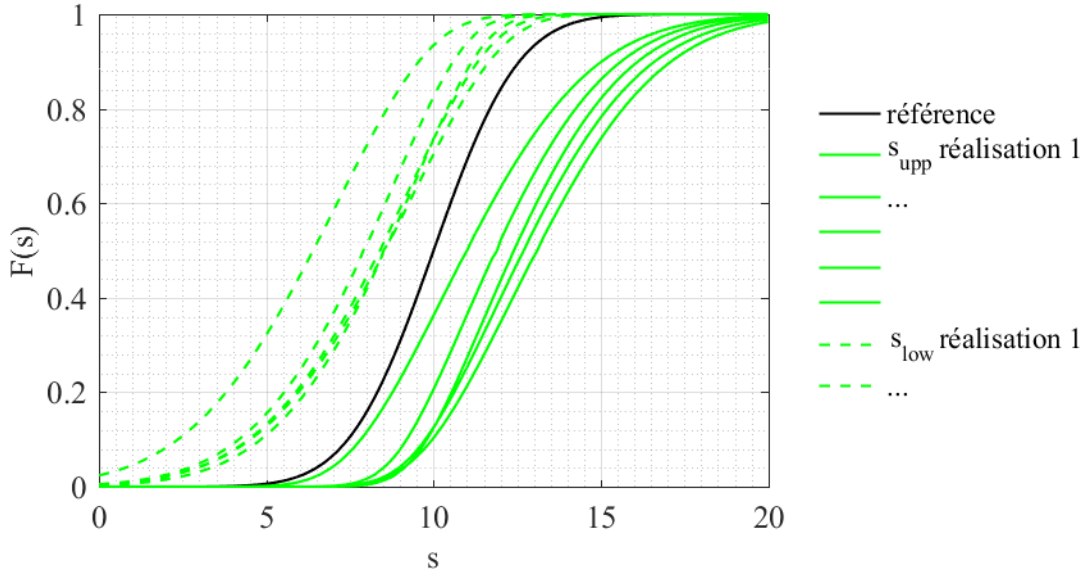

Figure 15 : Illustration du caractère aléatoire des bornes à l'aide de 5 réalisations différentes de taille n et des estimations de L et U correspondantes.

Lorsque les bornes s_{low} et s_{upp} sont calculées à partir de (16) avec un intervalle de confiance de 95 %, on s'attend à ce que le seuil ciblé se situe effectivement dans l'intervalle défini par les bornes inférieure et supérieure pour 95 échantillons sur 100. Autrement dit, le quantile à 97,5% de la borne inférieure ainsi que le quantile à 2,5% de la borne supérieure doivent être égaux au seuil ciblé. Le but de cette section est de vérifier cette attente, formalisée dans (17). Le seuil ciblé est $s_{5\%}$, i.e. le quantile $\alpha = 5\ \%$ de la susceptibilité. Pour vérifier cette attente, nous calculons la distribution des bornes à partir de 1000 échantillons. La distribution des bornes s_{low} et s_{upp} est représentée lorsque la taille des échantillons est de n=100 dans la Figure 16 (cas N) et dans la Figure 17 (cas W). Lorsque n=5, les résultats sont reportés en Figure 18 (cas N), et en Figure 19 (cas W).

Dans la Figure 16, ainsi que dans la Figure 17, la taille de l'échantillon est n=100, la distribution ne diffère pas selon la méthode d'inférence. De plus, le quantile à 2,5% de la borne inférieure et le quantile à 97,5% de la borne supérieure sont exactement égaux à $s_{5\%}$, comme attendu.

Dans la Figure 18, le quantile à 97,5% de la borne inférieure par l'IMV est très inférieur à $s_{5\%}$, alors que pour l'IB on obtient le résultat attendu. Pour la limite supérieure, le quantile à 2,5% de l'IB et de l'IMV sont proches de $s_{5\%}$. Par conséquent, la borne estimée par l'IMV présente une marge trop importante alors que l'estimation par l'IB est plus optimale.

Dans la Figure 19, le quantile à 97,5% de la borne inférieure par l'IB est inférieur au quantile $s_{5\%}$ alors que le quantile de l'IMV s'en approche. Les quantiles à 2,5% des bornes supérieures sont proche de $s_{5\%}$. Par conséquent, la borne estimée par l'IB présente une marge excessive alors que l'estimation fournie par l'IMV est plus optimale.

Les intervalles calculés par (16) sont donc validés. En effet, lorsque n=100, nous obtenons exactement ce qui est attendu. Lorsque n=5, la différence entre les attentes et les observations provient de la très faible taille d'échantillon.

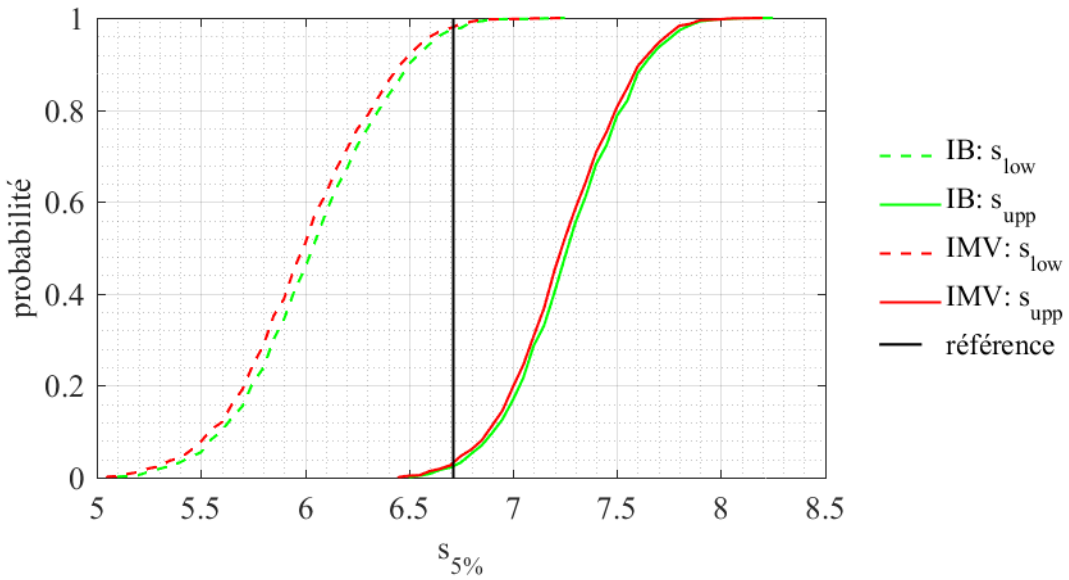

Figure 16 : Loi N, distributions des bornes s_{low} et s_{upp} avec un intervalle de confiance à 95 % encadrant le quantile 5 % avec une taille d'échantillon de n=100.

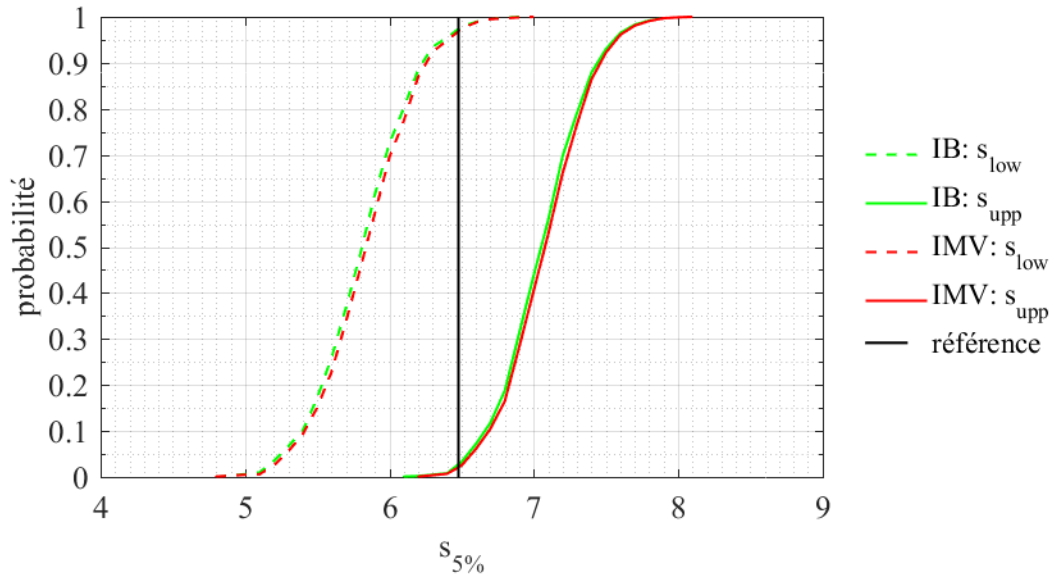

Figure 17 : Loi W, distributions des bornes s_{low} et s_{upp} avec un intervalle de confiance à 95 % encadrant le quantile 5 % avec une taille d'échantillon de n=100.

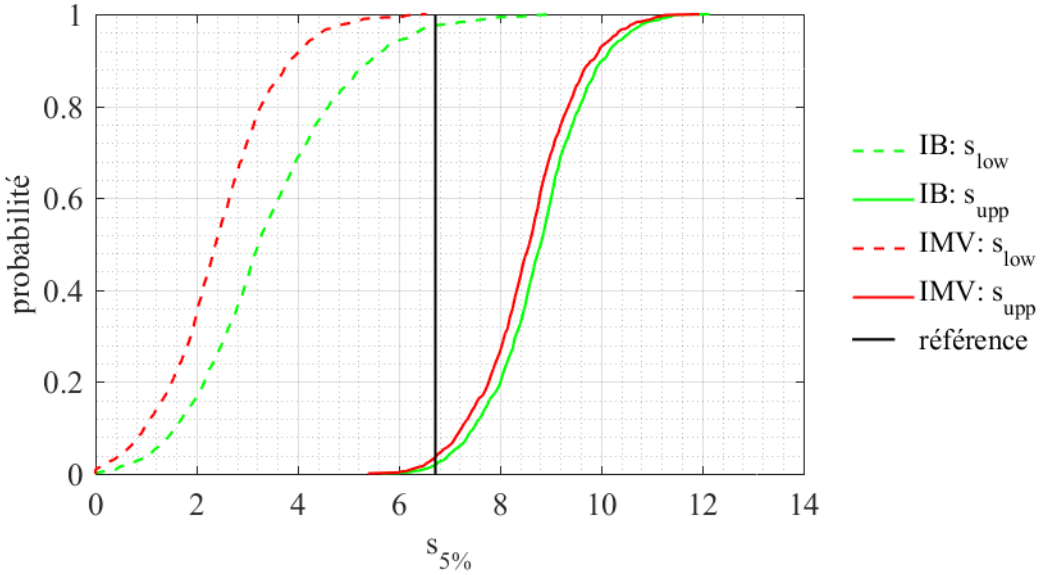

Figure 18 : Loi N distribution des bornes s_{low} et s_{upp} avec un intervalle de confiance à 95 % encadrant le quantile 5 % avec une taille d'échantillon de n=5.

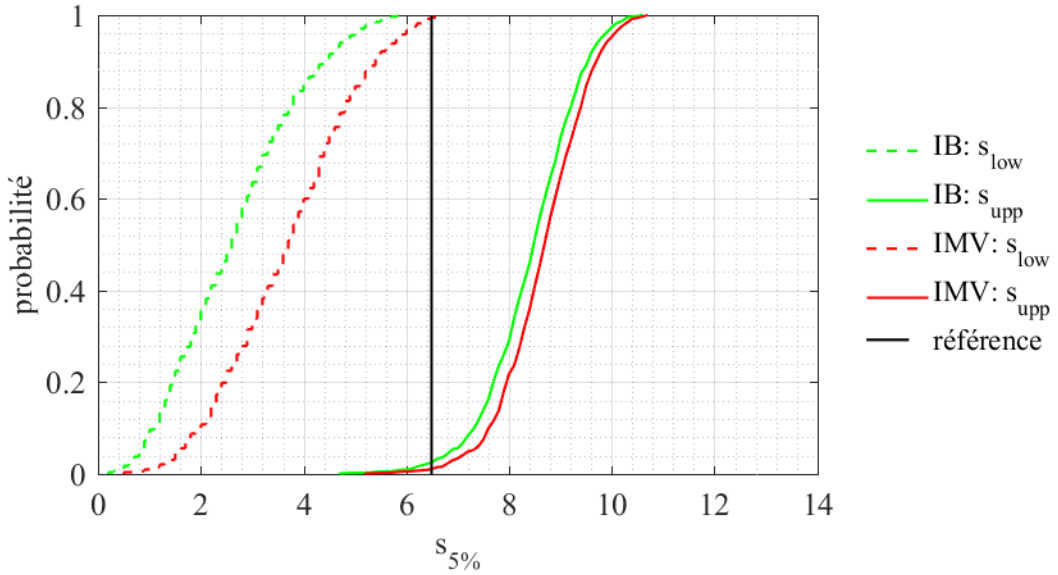

Figure 19 : Loi W, distributions des bornes s_{low} et s_{upp} avec un intervalle de confiance à 95 % encadrant le quantile 5 % avec une taille d'échantillon de n=5.

2. Encadrement dans le cas d'une distribution Normale

La procédure décrite au début de la section V est appliquée maintenant lorsque la distribution est N en faisant varier la taille de l'échantillon. Dans la Figure 20, nous traçons la moyenne des limites de s_α avec des barres d'erreur. La longueur d'une barre est égale à deux fois l'écart type de la borne calculée.

Dans le cas des échantillons de grande taille, les deux inférences mènent à des limites similaires. De plus, il n'est pas rentable d'augmenter la taille d'échantillon au-delà de *n*=40, car une augmentation beaucoup plus importante de *n* est nécessaire afin d'obtenir des limites significativement plus étroites.

Pour les petits échantillons, la différence entre les deux inférences est plus importante pour la

borne inférieure. Le minimum (moyenne moins écart-type) de la limite inférieure de l'IB est plus proche du minimum de la valeur réelle. Pour la limite supérieure, c'est le contraire. Par conséquent, l'IB est plus performante parce que les bornes sont légèrement plus étroites grâce à la limite inférieure.

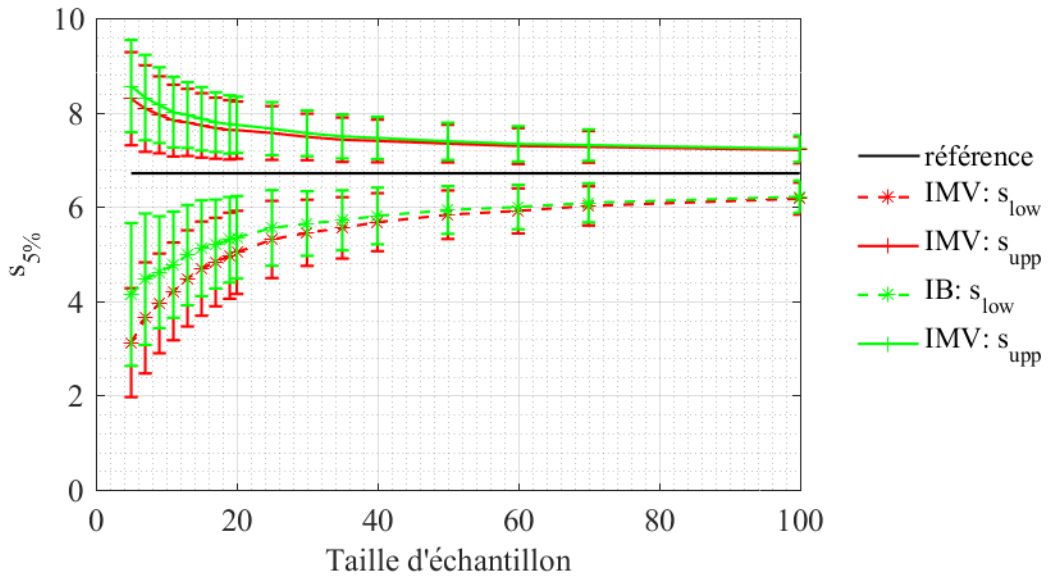

Figure 20 : Loi N, moyenne et écart type des bornes s_{low} et s_{upp} encadrant le quantile 5 % en fonction de la taille d'échantillon.

3. Encadrement dans le cas d'une distribution Weibull

La procédure décrite au début de la section V est appliquée maintenant lorsque la distribution est W. Dans la Figure 21, nous traçons la moyenne des limites de la susceptibilité avec des barres d'erreur (la longueur d'une barre est égale à deux fois l'écart-type de la limite calculée) à partir de l'IB et de l'IMV.

Comme dans le cas N, il n'est pas rentable d'augmenter la taille d'échantillon au-delà de 40 environ. Pour de plus petits échantillons, l'IMV est plus performante pour les mêmes raisons que l'IB l'est dans le cas N.

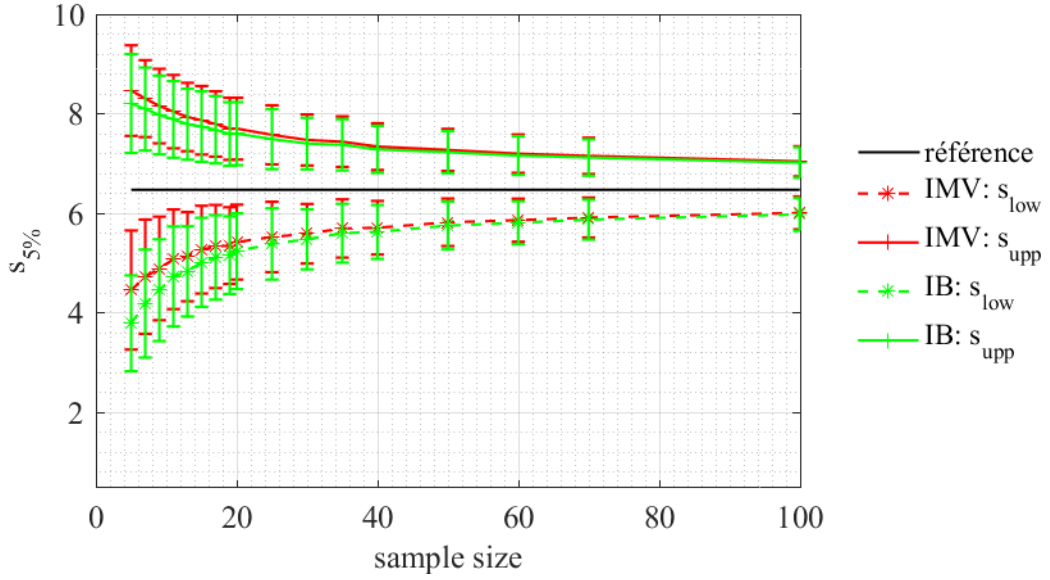

Figure 21 : Loi W, moyenne et écart type des bornes s_{low} et s_{upp} encadrant le quantile 5 % en fonction de la taille d'échantillon.

VI. Applications CEM

Dans les études précédentes (Sections III, IV et V), la génération des seuils était contrôlée puisque la distribution vraie était connue, quel que soit son type (N, LN ou W) et ses paramètres $\boldsymbol{\theta}$. Dans la pratique, ce n'est évidemment pas le cas. De plus, il est très difficile de tester un grand nombre d'équipements. Par conséquent, nous ne serons jamais en mesure d'obtenir la vraie distribution avec des équipements réels. C'est pourquoi nous proposons deux cas d'application : un exemple virtuel d'une part (sous-section 1) et un exemple de mesures sur des équipements réels d'autre part (sous-section 2).

1. Etude de la susceptibilité d'un équipement virtuel

a. Présentation du cas d'étude

Dans cette sous-section, la génération aléatoire de seuils de susceptibilité provient d'une simulation électromagnétique d'équipements électroniques virtuels. Il s'agit d'un cas intermédiaire entre la simulation théorique et un cas réel de mesures.

Dans ce cas d'étude, la variabilité du seuil de susceptibilité de l'équipement est assurée par une variation des paramètres physiques d'une simulation numérique. La fonction de distribution est inconnue, mais une pseudo-référence peut être tout de même obtenue. En effet, à condition que l'équipement virtuel soit rapide à simuler, nous pouvons simuler beaucoup de copies d'équipements. La simulation d'un grand échantillon nous permet d'obtenir une bonne estimation de la distribution réelle. Cette estimation sera considérée comme la pseudo-référence lorsque nous appliquerons les méthodes d'inférence sur de plus petits échantillons. Ces petits échantillons seront des sous-échantillons de celui de référence.

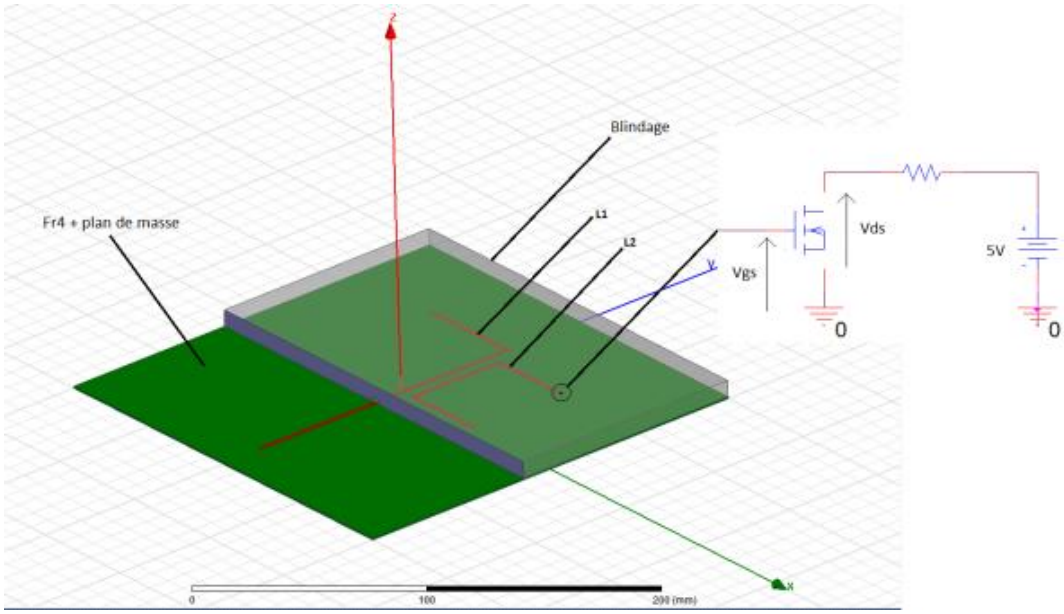

Figure 22 : Equipment électronique virtuel.

L'équipement est présenté à la Figure 22. Il s'agit d'une carte de circuit imprimé avec deux lignes micro-ruban *L1* et *L2*. *L1* se termine par une charge résistive alors que *L2* est terminée à une extrémité par une résistance et à l'autre par un transistor. *L2* est entièrement recouverte d'un blindage de protection tandis que *L1* ne l'est que partiellement. Une petite ouverture est pratiquée à l'interface du blindage pour que la ligne puisse passer à travers le blindage. La partie non couverte de *L1* est éclairée par une onde électromagnétique plane continue, à 500 MHz, d'incidence normale et polarisée dans la direction parallèle à la partie non couverte de *L1*. Sous le blindage, *L1* et *L2* sont proches l'une de l'autre, ce qui provoque un couplage par diaphonie. Les dimensions nominales de la conception sont : longueur de *L1*=205 *mm*, longueur de *L2*=130 *mm* ; le plan de masse a une surface de 250 x 250 *mm* ; l'épaisseur du substrat Fr4 est de 1,7 *mm.* La variabilité de l'équipement est assurée par le caractère aléatoire de multiples paramètres (longueur des lignes, charges, capacité parasite du transistor [52]) autour de la valeur nominale de conception (variabilité selon une loi uniforme avec une plage de variation de 10 % autour de la valeur nominale). Par conséquent, chaque copie d'équipement est une réalisation aléatoire.

L'intensité du champ est augmentée jusqu'à provoquer une défaillance. Le critère de défaillance est associé à la sortie (*Vds*) d'un transistor monté en inverseur. Lorsqu'aucune contrainte n'est appliquée (aucun champ incident), la tension d'entrée (*Vgs*) est à 0 *V* et donc la sortie (*Vds*) est à 5 *V*. Le transistor est ainsi bloqué. Si *Vds* est inférieur à une valeur arbitraire *Vth*=4 *V*, le transistor n'est plus considéré comme bloqué, ce que nous considérons comme une défaillance de l'équipement.

Nous avons simulé *nc*=2000 copies de cet équipement virtuel avec HFSS, formant ainsi l'échantillon de référence. Le nombre de réalisations *nc* a été choisi beaucoup plus grand que les échantillons de taille *n* d'intérêt (quelques dizaines), mais pas trop grand pour être compatible avec des temps de simulation raisonnables (environ 23h pour les 2000 simulations). Avec *Vth*=4 *V*, les seuils *nc* simulés sont représentés sous forme d'histogramme dans la Figure 23. Le choix de *Vth* n'a pas d'importance puisque sa variation n'implique qu'une translation sans déformation de la distribution des seuils. Par exemple, si *Vth* diminue, la distribution est translatée à des niveaux supérieurs. Dans la Figure 23, un ajustement (au sens IMV) avec les trois distributions est également tracé. L'ajustement LN est plus éloigné de la distribution empirique. Les distributions W et N semblent toutes deux bien correspondre (du moins visuellement) aux données. Le CV de l'échantillon de référence étant égal à 23.45 %, ceci est cohérent avec la Figure 8 où les distributions W et N sont similaires pour un CV proche de 30 %.

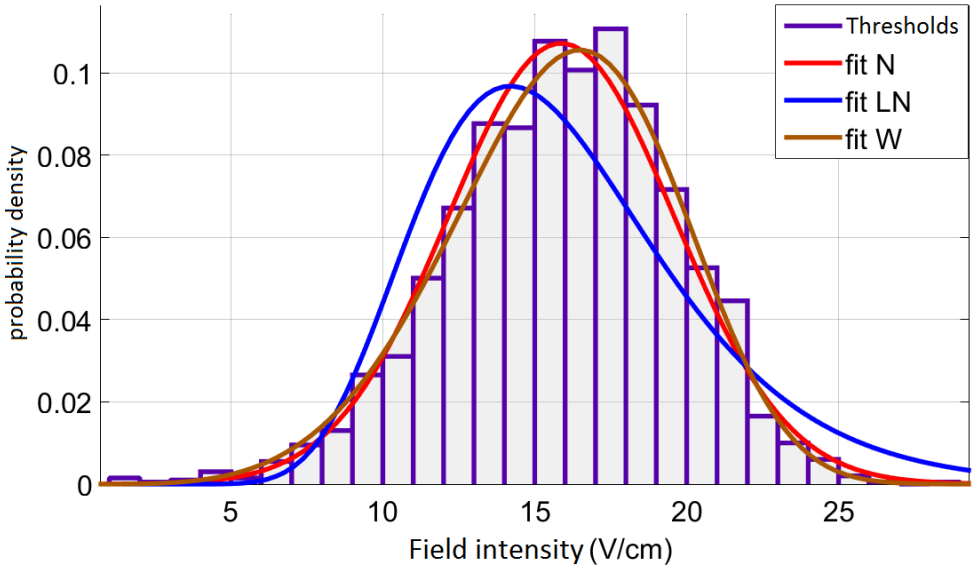

Figure 23 : Ensemble des seuils simulés constituant l'échantillon de référence et ajustements au sens IMV des lois N, LN et W.

À partir de l'échantillon de référence, trois quantiles $s_{5\%}$ ont été calculés. Le premier est évalué uniquement à partir de l'échantillon (quantile empirique). Le deuxième est estimé à partir de l'ajustement à la loi N aux données d'entrée tandis que le troisième provient de l'ajustement à la loi W. Il est difficile de décider laquelle de ces trois estimations de quantiles doit être considérée comme celle de référence. Cependant, la taille de l'échantillon de référence (2000) est très largement supérieure aux tailles qui seront testées (quelques dizaines). La variabilité des estimateurs issus de l'échantillon référence est donc largement inférieure à celle qui sera issue des petits échantillons.

b. Identification de la loi

Nous passons maintenant à l'étude visant à identifier la distribution. A chaque étape, nous avons sélectionné au hasard *n* équipements parmi *nc* équipements de l'échantillon de référence. Nous appliquons ensuite l'IB et l'IMV. Nous sélectionnons l'hypothèse correspondant au BIC minimum puis nous calculons les proportions de bonnes décisions. Comme le montre la Figure 23, les distributions W et N correspondent mieux aux données qu'une distribution LN. C'est pourquoi nous considérons qu'une bonne décision est de choisir indifféremment N ou W, tandis que choisir LN est une mauvaise décision.

Les résultats sont rassemblés dans le Tableau 8. Globalement, la prise de bonnes décisions est meilleure avec l'IB qu'avec l'IMV, car l'IB rejette plus souvent la distribution LN que l'IMV pour les trois tailles d'échantillon *n* testées. De plus, même si ce n'est pas déterminant ici, l'IB sélectionne plus souvent W que N alors que l'IMV sélectionne plus souvent N que W.

Tableau 8 : Proportion de prise de bonnes (vert)/mauvaises (rouge) décisions (%) concernant l'identification de la loi sous-jacente de seuil de susceptibilité pour l'équipement virtuel de la Figure 22.

	n **(→)** **Hypothèse choisie (↓)**	5	10	15
BIC(BI)	H_N	0.9	5.2	9.3
	H_{LN}	13.4	12.9	11.4
	H_W	85.7	81.9	79.3
BIC(IMV)	H_N	31.4	38	40.9
	H_{LN}	52.1	37.7	28.2
	H_W	16.5	24.3	30.9

c. Encadrement probabiliste du seuil de susceptibilité

Les bornes sont calculées cette fois-ci en choisissant l'hypothèse de distribution ayant le BIC minimum. Les deux sources d'erreurs dont nous parlions dès l'Introduction de ce chapitre, sont présentes dans ce cas d'étude. Ces erreurs concernent : le choix de l'hypothèse d'une part, et l'estimation de paramètres d'autre part.

Dans cette dernière sous-section, une étude Monte Carlo est menée afin d'estimer la distribution des bornes inférieures (encadrant s_{low}) et supérieures (encadrant s_{upp}) dans le cas particulier où nous ciblons le quantile à 5 % de la susceptibilité.

La moyenne et l'écart type des bornes sont fournis en Figure 24. Les bornes, calculées avec un intervalle de confiance de 95 %, convergent lentement vers la pseudo-référence. Les deux méthodes d'inférence ont des performances très similaires au-delà de n=30. Pour un échantillon de très petite taille, la moyenne de la limite inférieure de l'IB est plus proche de la référence. Le minimum, correspondant à la moyenne moins l'écart-type, pour la limite inférieure est presque identique pour les deux inférences. Pour les tailles extrêmement petites, la moyenne augmentée de l'écart-type est légèrement plus proche de la référence dans le cas de la limite inférieure de l'IB. En conséquence, l'IB offre un léger avantage sur l'IMV grâce à la borne inférieure.

Nous traçons, dans la Figure 25, la distribution des bornes encadrant $s_{5\%}$, pour un petit échantillon (n=5). Les bornes supérieures des deux inférences sont très semblables. Pour la borne inférieure, le quantile à 97,5 % fourni par l'IMV est plus proche du $s_{5\%}$ recherché que le quantile à 97,5 % estimé par l'IB. Nous remarquons tout de même qu'un risque de légère surestimation du seuil $s_{5\%}$ existe pour l'estimation IMV.

Quant à la Figure 26, elle présente l'estimation de la distribution des bornes encadrant $s_{5\%}$, pour un échantillon de bien plus grande taille (n=100). Les distributions des bornes des deux inférences sont confondues. De plus, la variance des bornes s'étant beaucoup réduite, la différence entre les trois méthodes d'estimation du seuil de référence (empirique, par ajustement N, puis par ajustement W) devient plus apparente. Pour les bornes supérieures, le quantile à 2,5% de s_{upp} est égal au quantile empirique. Par contre, s'agissant de la borne inférieure, la probabilité que s_{low} soit inférieure au quantile empirique n'est que de 93.5 % au lieu de 97.5 %. En effet, quantile à 97.5% de s_{low} est égal au quantile issu de l'ajustement W. Ceci met en évidence l'effet de la méconnaissance de la vraie nature de la distribution de la susceptibilité qui, potentiellement dans le cadre de ce cas test n'est ni exactement du type N ou du type W.

Il ressort de cette étude de cas que les méthodes d'inférence peuvent efficacement encadrer le

seuil de susceptibilité avec une précision (largeur des bornes) dépendante de la taille d'échantillon. La fiabilité de ces bornes est certes moins bonne que dans les cas théoriques mais celle-ci reste relativement proche de la valeur théorique attendue (erreur de 5 % pour les tailles très petites). La technique de Bootstrap peut être envisagée afin d'améliorer cette fiabilité. Cette technique est abordée comme piste de réflexion en conclusion de ce chapitre.

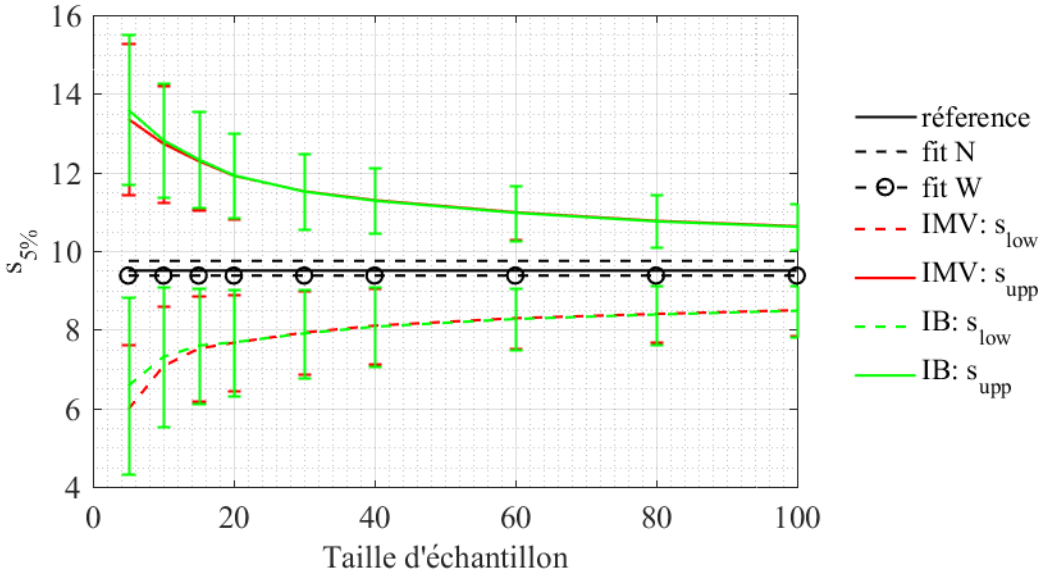

Figure 24 : Equipement virtuel. Encadrement probabiliste du quantile 5 % de la distribution de susceptibilité en fonction de la taille d'échantillon (s_{low} et s_{upp}).

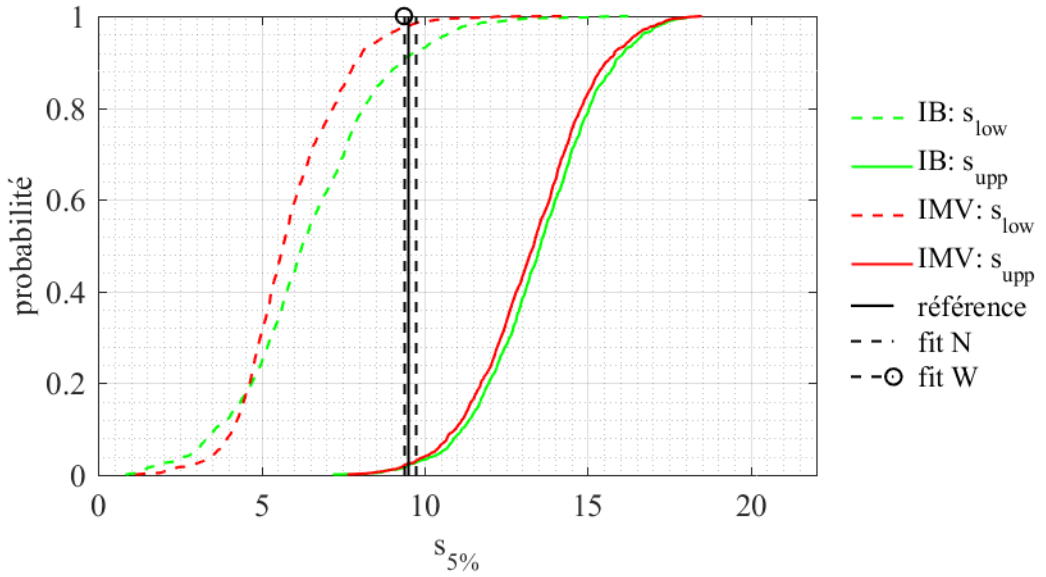

Figure 25 : Equipement virtuel. Échantillons de taille n=5, distributions des bornes (s_{low} et s_{upp}), calculées avec une confiance de 95 % encadrant le quantile 5 %.

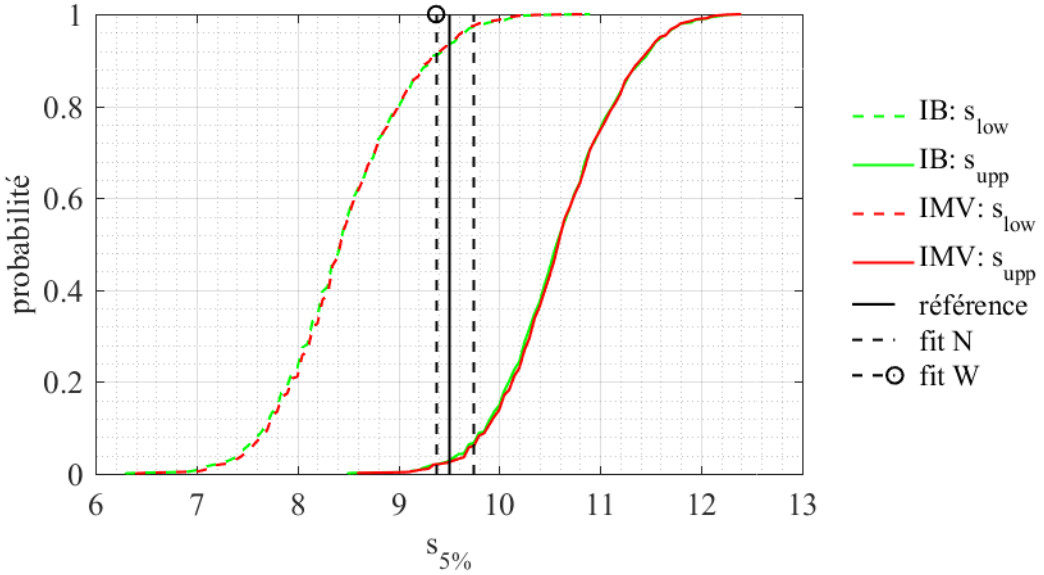

Figure 26 : Equipment virtuel. Échantillons de taille n=100, distributions des bornes (s_{low} et s_{upp}), calculées avec une confiance de 95 % encadrant le quantile 5 %.

2. Etude de la susceptibilité d'alimentations

a. Présentation du cas d'étude

Dans le cadre des études de susceptibilité d'équipements électroniques, le CEA de Gramat, a mesuré la susceptibilité de deux familles d'alimentations AC/DC lors d'injections de courant. Les seuils présentés ici ont été normalisés pour des raisons de confidentialité.

Aucune référence n'est disponible pour l'équipement testé, et nous ne disposons d'aucune connaissance préalable sur la distribution de la susceptibilité. Par conséquent, il n'est pas possible de comparer les résultats des inférences puisque nous n'avons pas la distribution vraie de la susceptibilité ni même de pseudo-référence comme dans le cas de l'équipement virtuel.

Les deux familles d'alimentations testées, sont nommées arbitrairement A et B avec une taille d'échantillon de 12 et 21 respectivement. Dans le Tableau 9 et le Tableau 10, nous présentons les données de mesures. Le niveau de contrainte ayant été discrétisé, les données sont des intervalles.

Tableau 9 : Seuils de défaillance mesurés (Ampères) lors de tests d'injection de courant sur des alimentations (Famille A).

Numero	1	2	3	4	5	6	7	8	9	10	11	12
Data	[4 , 5.6]	[4 , 5.6]	[5.6 , 7.3]	[7.3 , 8.9]	[7.3 , 8.9]	[7.3 , 8.9]	[7.3 , 8.9]	[7.3 , 8.9]	[13.7 , 16.9]	[8.9 , 10.5]	[8.9 , 10.5]	[12.1 , 13.7]

Tableau 10 : Seuils de défaillance mesurés (Ampères) lors de tests d'injection de courant sur des alimentations (Famille B).

Numero	1	2	3	4	5	6	7	8	9	10	11
Data	[1.8 , 3.9]	[3.9 , 6.3]	[6.3 , 9.0]	[12.1 , 14.7]	[6.3 , 9.0]	[3.9 , 6.3]	[1.8 , 3.9]	[6.3 , 9.0]	[12.1 , 14.7]	[1.8 , 3.9]	[12.1 , 14.7]
Numero	12	13	14	15	16	17	18	19	20	21	
Data	[3.9 , 6.3]	[1.8 , 3.39]	[6.3 , 9.0]	[9.0 , 12.1]	[1.8 , 3.9]	[9.0 , 12.1]	[1.8 , 3.9]	[9.0 , 12.1]	[17.9 , 21.4]	[1.8 , 3.9]	

b. Encadrement probabiliste du seuil de susceptibilité

Dans le Tableau 11, nous présentons le résultat de l'analyse des mesures. Le critère BIC est utilisé pour choisir un type de distribution. Au vu du critère BIC, la distribution retenue est de type LN. L'hypothèse choisie pour chaque famille est aussi rapportée. Enfin, nous présentons les limites de $s_{5\%}$ déterminées selon la procédure décrite dans ce chapitre.

Tableau 11 : Résultat de l'inférence du seuil de susceptibilité des deux familles d'alimentations.

Famille	**Choix d'hypothèse (BIC)**	**Bornes de $s_{5\%}$ (IMV)**	**Bornes de $s_{5\%}$ (IB)**
A	LN	[2.95 , 6.52]	[3.49, 6.69]
B	LN	[1.85 , 4.30]	[2.03, 4.42]

Dans la Figure 27, nous traçons le résultat pour la *cdf* empirique, la *cdf* la plus probable et les bornes de p_f à un niveau de confiance de 95 % (*U(s)* et *L(s)*) pour la famille d'alimentations A. Les bornes présentent des intervalles conséquents puisque la taille d'échantillon est faible. Il se trouve qu'avec cet échantillon, les bornes fournies par l'IMV et l'IB sont relativement semblables, ce qui confirme que le choix de l'une ou l'autre des méthodes d'inférence est peu discriminant.

Les bornes de $s_{5\%}$ présentent des intervalles conséquents. C'est tout à fait logique au regard de la taille de l'échantillon très limitée. Les bornes supérieures de l'IMV et IB sont relativement semblables tandis que la borne inférieure fournie par l'IMV est plus basse que celle de l'IB. Du point de vue d'une analyse de risque, l'estimation du pire cas sera privilégiée. Ainsi, la pire des estimations (IMV) indique que pour la famille A, le seuil de susceptibilité à 5 % est très probablement plus grand que 2.95 (1.85 pour la famille B).

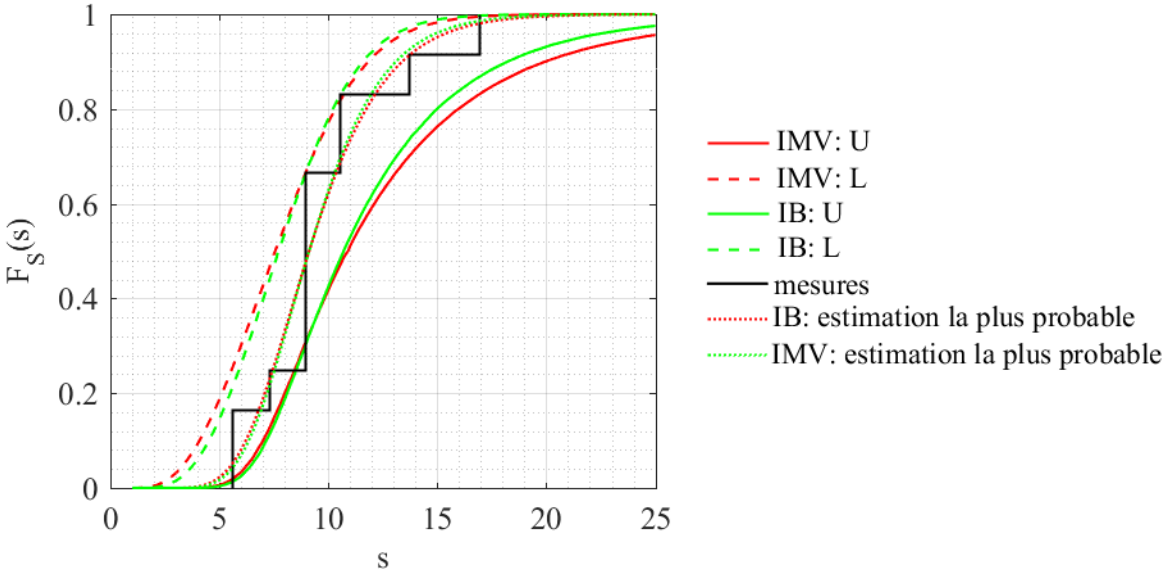

Figure 27 : Famille d'alimentations A : mesures, U(s) et L(s) ainsi que $p_f(s)$ le plus probable.

VII. Conclusion et pistes de réflexion

1. Conclusion

Ce chapitre a été consacré à l'inférence statistique de la distribution de susceptibilité d'équipements électroniques en prenant en compte le faible nombre d'équipements disponibles pour les essais. Nous avons comparé deux types d'inférence paramétrique : bayésienne et celle du maximum de vraisemblance, bien que cette dernière puisse être considérée comme un cas particulier de la première.

Quatre études ont été menées. Dans la première (section III), nous avons estimé la nature de la distribution, c'est-à-dire l'expression de la densité de probabilité ou de la fonction de répartition (*pdf/cdf*). Dans le cadre d'une étude Monte Carlo à partir de distributions théoriques connues a priori, nous avons appliqué les deux inférences afin de sélectionner une distribution parmi trois possibilités (N, LN, W). La deuxième étude (section IV) visait à estimer les paramètres de la distribution *pdf/cdf* lorsque celle-ci est connue a priori. Une troisième étude (section V) visait à encadrer par un intervalle de confiance la probabilité de défaillance de 5 % correspondant à la zone à risque en cas d'agression par une IEMI. Enfin, une quatrième étude (section VI) a été menée dont l'objectif était d'estimer la nature de la distribution, ses paramètres ainsi qu'un intervalle encadrant la probabilité de défaillance de 5 %.

Ces quatre études suggèrent qu'à partir de $n \approx 40$ équipements, la différence entre les deux méthodes d'inférence n'est pas significative. De plus, il n'est pas nécessairement rentable d'augmenter la taille d'échantillon au-delà. En effet, une augmentation beaucoup plus importante de n est nécessaire afin d'obtenir des limites significativement plus étroites. Autrement dit, la dérivée de la moyenne et de l'écart type de l'estimateur diminue rapidement lorsque n augmente. Pour de plus petits échantillons ($n \leq 40$), nous avons rassemblé dans le Tableau 12 le résultat de la comparaison entre l'IB et l'IMV pour tous les cas traités, à l'exception du cas où aucune référence n'est disponible (mesures des alimentations). Parmi les 11 comparaisons effectuées, l'IB n'est plus performante que l'IMV que dans 50 % des cas.

Tableau 12 : Bilan des comparaisons entre l'IB et l'IMV dans chacun des cas traités dans ce chapitre, uniquement pour de petites tailles d'échantillon ($n \leq 40$).

Etude (section)	Choix de la distribution (III)			Estimation des paramètres (IV)				Encadrement (V)		Equipement virtuel (VI)	
Sous cas	N	LN	W	N		W		N	W	Choix distribution	Encadrement
				θ_1	θ_2	θ_1	θ_2				
Renvoi	Figure 9	Figure 10		Figure 11	Figure 12	Figure 13	Figure 14	Figure 20	Figure 21	Tableau 8	Figure 24
IB			✓	=	✓*v*		✓*b*	✓		✓	✓
IMV	✓	✓			✓*b*	✓	✓*v*		✓		

Avantage : ✓ , stricte équivalence : =, avantage si la réduction du biais (variance) est privilégiée au détriment de la variance (biais) : ✓ *b* (*v*).

Il n'a donc pas été possible d'identifier clairement la méthode d'inférence à privilégier. Si un choix de méthode n'est pas possible, il est en revanche beaucoup plus important de ne pas se contenter d'une estimation la plus probable. En effet, le risque est de négliger à tort, la grande incertitude due à la petite taille de l'échantillon. A ce titre, l'IMV tout comme l'IB fournit un intervalle de confiance des paramètres de la distribution. L'approche pertinente pour une analyse de risque consiste donc à utiliser ces intervalles pour encadrer la probabilité de défaillance, que cette estimation provienne d'une IB ou d'une IMV.

2. Pistes de réflexion

Dans ce chapitre, nous avons utilisé des a priori non informatifs pour l'inférence bayésienne. Si des connaissances préalables sont disponibles (expériences antérieures, avis d'experts...), l'inférence bayésienne peut en tenir compte et très probablement améliorer les performances.

Nous avons supposé que la distribution sous-jacente du seuil n'était pas tronquée. Une distribution tronquée est le résultat de la restriction du domaine des valeurs possibles. Cependant, cela ne change pas la procédure d'inférence. Il suffit de modifier l'expression de la *pdf/cdf* supposée sous-jacente et rajouter les bornes de troncature comme composantes inconnues à $\boldsymbol{\theta}$. En effet, si S_T est une variable aléatoire issue de la troncature de S entre A et Ω tel que $S_T = S.\mathbb{I}_{S\in]A,\Omega]}$, alors la *pdf* de T est f_T tel que:

$$f_{S_T}(s) = \frac{f_S(s)}{F_S(\Omega\ ) - F_S(A)}\mathbb{I}_{s\in]A,\Omega]}. \tag{23}$$

Les niveaux de test ont été considérés dans ce chapitre comme totalement maitrisés. Il n'y a donc pas d'incertitudes sur les valeurs collectées dans le vecteur $\boldsymbol{Data}$ définie dans (1) même dans le cas où $\delta_e > 0$. Afin de prendre en compte ces incertitudes, on pourrait considérer que ce n'est pas seulement S que l'on mesure mais S_B=S+B, où B serait une variable aléatoire associée au bruit de mesure dont la *pdf* est f_B. Dans ce cas, la *pdf* de S_B notée f_{S_B} est le produit de convolution des *pdf* de S et de B. Comme dans le cas de la troncature, il faut aussi rajouter les paramètres de f_B dans $\boldsymbol{\theta}$ comme composantes à déterminer. Selon les lois pour S et B choisie, la fonction f_{S_B} n'est pas forcément calculable analytiquement.

Nous avons traité uniquement des méthodes d'inférence paramétrique qui supposent un type de loi dont une expression analytique de la *pdf/cdf* est connue. Il existe cependant une autre famille

d'inférence, dite « non-paramétrique » [44], qui se passe de l'hypothèse de loi pour estimer directement des statistiques de la population. C'est le cas de l'estimateur de Kaplan-Meier [53] dont la variance peut être approximée par la formule de Greenwood afin de calculer les intervalles de confiance encadrant la probabilité de défaillance.

D'autres techniques non-paramétriques comme le Bootstrap [54] se basent sur des ré-échantillonnages multiples de l'échantillon de départ. Chacun des ré-échantillonnages est un tirage aléatoire avec remise (selon une loi uniforme discrète) de réalisations dans l'échantillon initial. On procède classiquement à autant de tirages qu'il y a de réalisations dans l'échantillon initial. On obtient ainsi un premier nouvel échantillon. De ce nouvel échantillon, on peut estimer la statistique d'intérêt (e.g. la moyenne, l'écart type,…). Dans le cadre de l'étude de la distribution de la susceptibilité, la statistique d'intérêt serait les quantiles, c'est-à-dire une estimation empirique de la *cdf* noté $\hat{F}(s)$ dans Figure 28. Il est possible d'appliquer un ajustement par noyau (classiquement gaussien) afin de lisser la distribution empirique. En répétant cette procédure on obtient un échantillon de la statistique d'intérêt. Finalement, on obtient un faisceau de courbes, permettant de réaliser le même type d'encadrements que ceux obtenus dans ce chapitre. L'avantage est que nous n'avons pas besoin de supposer une loi sous-jacente. Cependant, cette technique est forcément dépendante de l'échantillon de départ. Il conviendrait donc de procéder au même type d'étude (i.e. Monte Carlo) et de comparer les estimations fournies par le Bootstrap par rapports aux inférences traitées dans ce chapitre. D'après une étude préliminaire, il semblerait que l'IMV soit identique au Bootstrap lorsque la loi est connue. Cependant, le Bootstrap pourrait s'avérer avantageux dans un cas similaire aux cas d'applications de la section VI de ce chapitre, où la loi sous-jacente est à choisir.

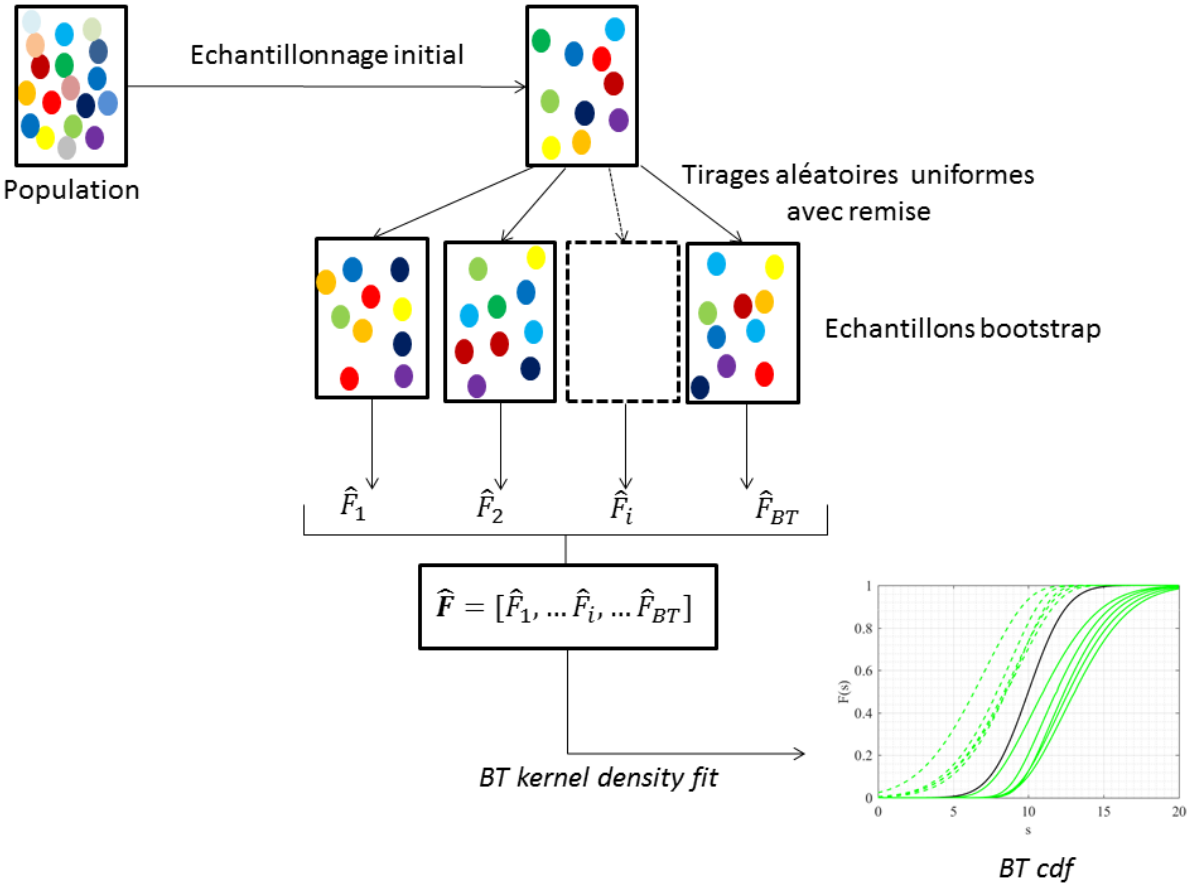

Figure 28 : Procédure d'estimation de la distribution de susceptibilité par la méthode de Bootstrap.

CHAPITRE III. RECHERCHE DE VALEURS EXTREMES DE CONTRAINTE PAR KRIGEAGE ET STRATIFICATION CONTROLEE

I. Introduction

Lors de l'agression électromagnétique d'un équipement par une IEMI, l'évaluation des risques à travers l'estimation de la probabilité de défaillance nécessite d'étudier le chemin de couplage de l'onde incidence jusqu'à l'entrée de l'équipement. Cependant, il existe de multiples incertitudes épistémiques, à la fois propres aux caractéristiques de l'onde (e.g. angle d'arrivée, intensité, polarisation, allure temporelle…), et aux chemins de couplage (couplage direct ou indirect de l'onde sur une entrée de l'équipement). Ces incertitudes concernent donc les paramètres dits « d'entrée » qui sont mieux décrits par des variables aléatoires. Ces variables aléatoires peuvent prendre plusieurs valeurs possibles décrites par des lois de probabilité reflétant notre degré d'incertitudes à leur sujet. Puisque les variables d'entrée peuvent prendre plusieurs valeurs, la « sortie », i.e. la grandeur d'intérêt (e.g. un courant, une tension) caractérisant le niveau de contrainte à l'entrée de l'équipement risque aussi de prendre plusieurs valeurs possibles, elle est donc également incertaine. Du fait de ce phénomène de propagation d'incertitudes, la contrainte exercée sur l'équipement est définie comme une variable aléatoire, tout comme l'était la susceptibilité.

Lors de l'étude de la susceptibilité, nous ne disposions d'aucun modèle de l'équipement car aucune connaissance de la structure de l'équipement ou aucun modèle de ce dernier n'était disponible. Ainsi, l'étude menée au chapitre précédent s'est basée exclusivement sur des mesures afin d'évaluer la distribution probabiliste du niveau de susceptibilité. En revanche, dans le cadre de l'étude de la contrainte exercée sur l'équipement, nous supposons disposer d'un modèle électromagnétique fiable du chemin de couplage. La description paramétrique de ce modèle est liée à la connaissance de la source d'agression, des voies de couplage vers le système hébergeant l'équipement ciblé par l'analyse et dont la localisation et les voies d'accès (filaires ou non) sont également décrites. Une modélisation de l'environnement de l'équipement et donc du chemin de couplage est possible.

La modélisation de l'agression électromagnétique requiert la résolution numérique des équations de Maxwell, selon différentes méthodes. Ces méthodes sont le plus souvent de type *full-wave* (e.g. méthode des moments, différences finies…) et fournies par des logiciels commerciaux (e.g. HFSS, CST, FEKO,…) ou internes. Le CEA développe et exploite notamment le logiciel GORF 3D [55][56] basé sur la méthode des différences finies dans le domaine temporel (FDTD). Ces modèles numériques sont déterministes et nous allons les considérer exacts car leur erreur est négligeable devant la variabilité de la sortie. Malheureusement, ces modèles peuvent être coûteux en temps de calcul, particulièrement lorsque la géométrie 3D du système étudié est complexe et qu'il possède une dimension significative devant la longueur d'onde. De plus, nous cherchons à déterminer des valeurs de contraintes extrêmes, ce qui exige le recours à de nombreuses simulations. En effet, les valeurs extrêmes sont des réalisations de sortie rares, i.e. peu probables. Dans le cadre d'une étude de risque de défaillance d'un équipement face à une agression électromagnétique de type IEMI, ce sont ces réalisations qui permettent d'estimer la probabilité de défaillance. C'est pourquoi, nous nous focalisons sur la recherche de ces valeurs.

Cette recherche revient à estimer un quantile extrême de la distribution de la contrainte. Les quantiles permettent de décrire la répartition d'une variable aléatoire. Plus précisément, à un quantile associé à une probabilité α ($0<\alpha<1$) correspond une valeur de la variable aléatoire. Une quelconque réalisation de celle-ci possède alors une probabilité α de ne pas dépasser cette valeur et $1-\alpha$ de l'excéder.

L'accès à une estimation d'un quantile extrême et/ou à une distribution de valeurs extrêmes est particulièrement ardu et nécessite l'utilisation d'outils statistiques spécifiques permettant d'accélérer la convergence de l'estimation. Dans cette thèse, nous faisons appel à une technique appelée

Stratification Contrôlée (SC) qui est dédiée à la recherche de valeurs extrêmes. Cette technique s'efforce d'échantillonner judicieusement l'espace des variables d'entrée de manière à générer un plus grand échantillon de valeurs extrêmes. Cette technique nécessite un modèle de substitution aussi appelé « métamodèle » qui sera abrégé SM pour *Surrogate Model* en anglais. Un SM lui permet de contrôler cet échantillonnage afin d'être en mesure de sélectionner des réalisations d'entrée provoquant des valeurs extrêmes de sortie. L'association d'un SM et de la stratification contrôlée a pour ambition d'être beaucoup plus efficace que la technique d'estimation standard (utilisée couramment dans les sciences de l'ingénieur), qu'est l'approche empirique de la simulation Monte-Carlo (MCS).

Cependant, la question relative à la recherche et à la construction d'un bon SM est laissée jusqu'à présent totalement ouverte, en dehors des situations où un modèle physique simplifié pourrait être utilisé. Afin de répondre à cette question, nous apportons 2 contributions majeures. La première est l'introduction de la technique de stratification contrôlée couplée à un SM dont la construction est réalisée de manière systématique. La deuxième est la proposition d'un algorithme implémentant cette technique, tout en maitrisant le budget de simulation.

Ce chapitre est organisé en 5 sections, hormis l'introduction (section I) et la conclusion et pistes de réflexions (section VII). Tout d'abord, un état de l'art (section II) décrit l'estimation de quantiles à partir de la technique de la stratification contrôlée ainsi que par des SMs. Ensuite (section III) nous présentons 2 modèles représentatifs de problèmes CEM, qui seront ensuite utilisés tout au long du chapitre. Dans la section suivante (section IV), nous détaillons une première étude qui vise à comparer plusieurs SMs, afin de choisir le plus adapté à l'hybridation avec la méthode de la SC. Cette étude conclura que le krigeage est le meilleur candidat. Une seconde étude (section V) vise à évaluer l'apport de la méthode de la SC par référence à l'utilisation du seul krigeage. Enfin la dernière étude (section VI) propose une implémentation pratique de la K-SC sur un cas-test final représentatif d'un problème de CEM.

II. Etat de l'art

La plupart des phénomènes physiques, comme la propagation des ondes électromagnétiques, peuvent être modélisés. Un modèle est une fonction M qui permet de prédire une variable de sortie Y étant donné un vecteur de variables d'entrée $\boldsymbol{X}$:

$$Y = M(\boldsymbol{X}). \tag{24}$$

Les modèles M considérés ici sont une description exacte du phénomène physique déterministe mais sont coûteux en temps de calcul comme les simulations de type *full-wave* où les équations de Maxwell sont résolues numériquement. En raison des incertitudes épistémiques, l'entrée $\boldsymbol{X}$ est incertaine et donc considérée comme une variable aléatoire multivariée. L'incertitude se propage dans le modèle. Par conséquent, la variable Y est également incertaine et donc définie comme une variable aléatoire (ici univariée mais la généralisation au cas multivarié ne pose pas de difficultés).

L'objectif d'une étude de risque est de calculer un quantile y_α de la fonction de distribution cumulative (*cdf*) de la sortie Y. Par définition, le quantile y_α est lié à la *cdf* puisque :

$$F_Y(q_\alpha) = \mathbb{P}\{Y \leq y_\alpha\} = \alpha. \tag{25}$$

Un quantile de probabilité α est donc la valeur de la variable de sortie pour laquelle sa *cdf* atteint le niveau de probabilité α. Toute estimation de quantile fait appel à la construction d'un plan d'expérience (DOE : Design Of Experiment) [57]. Le DOE comprend, d'une part, des réalisations aléatoires des entrées et, d'autre part, les sorties correspondantes calculées avec le modèle M. La défaillance d'un équipement soumis à une IEMI survient lorsque la contrainte est supérieure à sa susceptibilité. Ainsi, les réalisations de la variable aléatoire de la contrainte ayant un impact sur la

probabilité de défaillance se situent dans la queue supérieure de la distribution. C'est pourquoi nous avons cherché à estimer des quantiles extrêmes de la distribution de la contrainte.

D'une manière générale, l'estimation d'un quantile peut se résumer aux trois étapes principales suivantes :

1. Échantillonner n réalisations de l'entrée $\boldsymbol{X}$ multivariée. A partir de ces réalisations, utiliser le modèle pour calculer les n sorties à l'aide du modèle M, i.e. n réalisations de la contrainte. Cette première étape est donc la construction d'un DOE. Afin d'observer au mieux l'ensemble des valeurs possibles, la qualité de l'échantillonnage est importante. C'est pourquoi nous avons choisi l'échantillonnage de type hypercube latin (LHS – latin hypercube sampling) qui consiste à subdiviser l'espace en multiples hypercubes identiques pour une répartition plus uniforme des réalisations [58].
2. Estimer la *cdf* de la sortie F_Y. Nous avons considéré trois méthodes possibles, issues de la littérature statistique :
    a. Estimation empirique.
    b. Utilisation de l'aptitude prédictive des SMs.
    c. Stratification contrôlée basée sur un SM.
3. Estimer le quantile visé, ce qui est simple une fois la deuxième étape franchie. En effet :

$$y_\alpha = F_Y^{-1}(\alpha) \tag{26}$$

Ces dernières années, certains auteurs ont tiré parti de concepts statistiques avancés pour une analyse de risque en présence d'incertitudes. La plupart de ces techniques sont basées sur une définition adéquate du DOE pour construire un SM [59] qui peut prédire de nombreuses sorties à faible coût de calcul, et finalement toutes statistiques d'intérêt. Différentes techniques stochastiques ont été appliquées avec succès afin de donner une vision plus réaliste des simulations électromagnétiques incluant les incertitudes. Parmi elles, les méthodes de transformation dites « unscented transforms » [60], de collocation stochastique basée sur les polynômes de Lagrange [61][62], la technique de krigeage [63][64][65][66][67], l'expansion du chaos polynomial [68][69][70][71] ont été proposées.

Les techniques citées précédemment ont été principalement appliquées à des simulations numériques incluant diverses applications : efficacité de blindage [72], diffusion et propagation [73], et/ou problèmes de CEM. Différentes techniques ont également été mises à profit dans un contexte CEM plus particulièrement dédiées à l'estimation des quantiles extrêmes, fondées sur des approches fiabilistes [74][75][76][77] ainsi que sur les méthodes dites de stratification contrôlée (SC) [78][79].

1. Estimation empirique d'un quantile par simulation de Monte Carlo (MCS)

La simulation par Monte Carlo (MCS) est l'approche standard pour estimer la distribution de sortie. Les estimateurs fournis par MCS sont appelés « Estimateurs Empiriques » (EE).

Notons qu'implicitement, la MCS utilise le modèle M qui est exact mais coûteux. Il est possible de faire de la MCS avec n'importe quel modèle de substitution ou SM. Ainsi, afin d'éviter une confusion, l'abréviation MCS est exclusivement dédiée à une utilisation du modèle M. Dans les cas où le modèle n'est pas M, l'abréviation ne sera pas utilisée au profit d'une version en toute lettre précisant le modèle utilisé si nécessaire.

Afin de réaliser une simulation de type MCS, un DOE est tout d'abord construit. Le DOE consiste en la définition d'un échantillon de réalisation du vecteur des variables d'entrée de taille n. Il s'agit typiquement d'un grand échantillon (n > 1000) DOE constitué typiquement par un échantillonnage selon la technique d'échantillonnage uniforme de type « hypercube latin » (LHS). Le modèle M est utilisé pour chaque réalisation afin de calculer la distribution empirique de la contrainte. Il en résulte un échantillon de sortie qui est alors ordonné pour fournir une estimation empirique de la

cdf. Enfin, une interpolation, généralement linéaire, est effectuée. Dans les résultats que nous présenterons plus loin dans ce chapitre, nous avons utilisé la fonction Matlab *quantile*, mettant en œuvre cette technique. Les techniques de rééchantillonnage [80] peuvent également affiner cette estimation.

Cette approche est simple et très robuste mais possède une convergence très lente, surtout si des valeurs extrêmes sont visées. La taille du DOE est donc généralement très grande, ce qui implique un temps de calcul important.

2. Estimation d'un quantile par métamodèle

a. Définition et métamodèles choisis

Un SM est un terme très général englobant toute fonction M' approximant le modèle M qui est exact mais coûteux :

$$Y \approx \hat{Y} = M'(\boldsymbol{X}) \tag{27}$$

Il existe trois types de SMs que nous qualifierons respectivement de « physique », « numérique » et « d'interpolation ». Tous ont pour ambition d'approximer le modèle M tout en étant nettement plus économiques en temps de calcul.

Un SM « physique » remplace la modélisation du phénomène physique complexe par la modélisation simplifiée, mais suffisamment représentative du phénomène observé. Par exemple, afin d'étudier l'adaptation d'une antenne, une simulation HFSS de type *full-wave* (modèle exact) du coefficient de réflexion d'une antenne pourrait être remplacée par un circuit résonant RLC. Les valeurs de R, L et C seraient à définir en fonction de la géométrie et des caractéristiques des matériaux utilisés. Pour construire un SM « physique », une bonne connaissance des phénomènes physiques impliqués est nécessaire. De tels SMs sont donc spécifiques à un problème donné et sont souvent limités par les hypothèses limitant leur validité (bande de fréquence, température, etc.). Par ailleurs, il est souvent très difficile d'obtenir un modèle physique suffisamment fiable de manière systématique sauf à tolérer un biais d'estimation parfois important et difficile à quantifier. Nous n'avons donc pas choisi cette approche.

Des SMs qualifiés de « numérique » auraient pu aussi être considérés. Par exemple, si le modèle *full-wave* avec un maillage très fin est un modèle exact de la réalité (i.e. erreur négligeable mais temps de calcul important), alors le même modèle mais avec un maillage plus lâche est un SM « numérique » (erreur plus importante mais temps de calcul accéléré). Ce choix de SM est pertinent et a déjà été traité dans [79]. Cependant, ce type de SM reste limité aux modèles se basant sur un maillage. De plus les propriétés obtenues ne permettent pas de garantir systématiquement une corrélation suffisante avec le modèle initial. Cette stratégie possède donc un champ d'applications limité et n'est donc pas retenue.

Le troisième type de SM repose sur des schémas d'interpolation. C'est le type le plus universel puisque ces SMs ont une vision de type boîte noire du modèle exact. En effet, ces SMs n'ont besoin que des réalisations d'entrées et des réalisations de sortie associées, sans savoir comment ces sorties sont calculées par le modèle exact. Dans cette famille nous pouvons distinguer encore plusieurs sous-familles de SMs : interpolations déterministes, stochastiques ou issues du *machine learning*. La sortie étant une variable aléatoire, une approche de type interpolation déterministe (splines,…) n'est pas appropriée. C'est pourquoi nous allons étudier deux types d'interpolation stochastique et deux algorithmes d'apprentissage de *machine learning*.

Les cinq SMs choisis sont : le krigeage (KRI) [63], le chaos polynomial ou *polynomial chaos expansion* (PCE) [81] , les machines à support vectoriel ou *Support Vector Machine* (SVM) [82], le réseau de neurones (*Neural Network*, NN en abrégé) [83] et enfin la combinaison du PCE avec le

krigeage (PCE_KRI) [84]. Ce choix repose sur le fait que ces algorithmes sont déjà implémentés dans au moins un des langages de programmation les plus utilisés par la communauté CEM, à savoir Matlab, Python et R. De plus, choisir plusieurs SMs au lieu de se contenter d'un seul est motivé par le fait qu'il est impossible de déterminer a priori celui qui sera le plus performant pour un modèle donné. Cependant, il est possible de les comparer entre eux à partir de DOEs pour un coût de calcul limité. Un choix pertinent de SM est dès lors possible.

b. Métamodèle pour la recherche d'événements extrêmes

Une fois qu'un SM est construit à partir d'un DOE de taille *n* (typiquement quelques dizaines ou centaines), le SM peut prédire un nombre quasi illimité de réalisations de sortie à faible coût. La propagation d'incertitudes à travers un SM est donc simple et rapide. Par exemple, nous pouvons prédire 1 million de sorties en quelques secondes puis estimer le quantile empirique de ces 1 millions de sorties. Si le SM est construit avec une taille de DOE suffisamment grande pour que sa prédiction soit précise, alors le quantile estimé le sera également aux erreurs résiduelles près du métamodèle.

Les différents SMs présentés plus haut ne sont pas spécifiquement dédiés à cibler la queue de distribution de la variable de sortie. Cependant, nous émettons l'hypothèse qu'ils peuvent être avantageusement combinés avec d'autres techniques dédiées aux valeurs extrêmes comme la stratification contrôlée (SC) [78], la simulation de sous-ensemble [85], l'échantillonnage par importance [86]. A l'exception de la stratification contrôlée, les techniques citées peuvent se passer de SM. En effet, la SC nécessite un modèle corrélé simple. Ce modèle corrélé simple peut être un SM de type « interpolation » comme nous le proposons. Dans [79], la SC a été appliquée à un problème de diaphonie avec un SM de type « numérique ». Nous pouvons également citer [87], où la SC est appliquée à un problème de mécanique.

c. Comparaisons ente les métamodèles en CEM

Dans la littérature, certains SMs ont été comparés entre eux : un PCE par rapport au SVM [88], un KRI par rapport au NN [89][90], un PCE par rapport à la collocation stochastique [91], un PCE par rapport au KRI [92].

Cependant, aucune comparaison de cinq SMs en même temps n'a été faite dans le cadre de la recherche d'événements extrêmes en CEM. Le problème est qu'une hiérarchie entre les SMs obtenue pour un modèle particulier ne peut prétendre être valable pour tous les modèles possibles. De plus, chaque SM est plus ou moins influencé par ses paramètres. En effet, chaque SM est disponible en plusieurs variantes selon le choix qui est fait lors de sa construction pour un paramètre particulier. Un ajustement est nécessaire afin de s'adapter au mieux aux modèles utilisés en section 3 de ce chapitre. Nous avons cependant observé pour nos modèles, que cet ajustement est moins crucial pour les méthodes KRI, du PCE et du SVM contrairement au NN.

La section suivante contient pour chacun des cinq SMs d'interpolation choisis : un bref rappel des bases mathématiques, des exemples de leur utilisation dans un contexte CEM, et les différents paramètres choisis.

3. Support Vector Machine (SVM)

Le SVM est à l'origine une technique servant de classificateur multi classes. Cependant, il peut être utilisé à des fins de régression. Dans ce but, SVM permet de trouver le meilleur hyperplan de régression [82]. En deux dimensions, il s'agit d'une régression linéaire. Grâce aux fonctions appelées « noyaux » qui transforment l'espace des variables initiales, il est toujours possible de retomber dans le cas d'une régression par hyperplan. La Figure 29 illustre une régression en deux dimensions par SVM dans le cas où les points sont ceux d'origine ou après avoir subi la transformation par un noyau. Dans le cadre d'une régression classique, on cherche à trouver une droite qui minimise la distance de chaque point par rapport à cette droite. Avec un SVM, on se focalise sur les distances des points en dehors

d'une zone de tolérance qui s'appelle un « tube » dans la terminologie propre aux SVMs. Cette zone correspond en 2D à l'espace entre les deux droites qui sont distante de +- ϵ de la droite de régression.

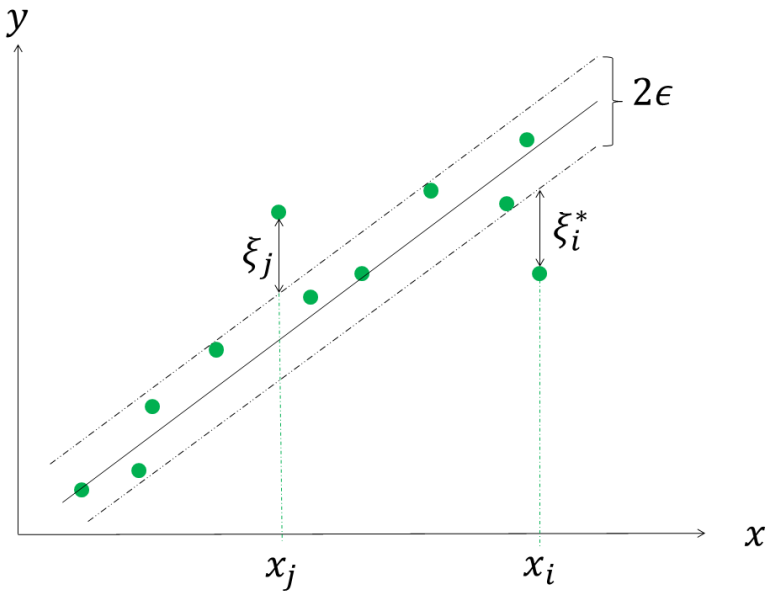

Figure 29 : Exemple d'utilisation d'un SVM dans le cadre d'une régression linéaire. Seuls les points x_j et x_i, qui sont en dehors du tube (défini ici par les deux droites pointillées noires) sont pénalisés.

Le calcul d'une régression par SVM consiste à trouver des poids *w* et scalaire *b* qui minimise :

$$\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n} \xi_i + \xi_i^* , \tag{28}$$

sous la contrainte :

$$\begin{cases} y_i - wx_i - b \leq \epsilon + \xi_i \\ wx_i + b - y_i \leq \epsilon + \xi_i^* \\ \xi_i^*, \xi \geq 0 \end{cases} \tag{29}$$

Un SVM est paramétré par la sanction pour violation de contrainte *C* qui est mesurée par ξ et ξ* ainsi que la tolérance ϵ du tube. De plus, quelle que soit la fonction noyau choisie, celle-ci est aussi paramétrique.

Le choix des paramètres (C et ϵ) est étudié dans [93][94][95], celui du kernel dans [96]. Une association du SVM avec la méthode fiabiliste de *subset simulation* a été réalisée dans [97]. En électromagnétisme, le SVM a été utilisé en micro-onde dans [98]. De plus, dans [88] un SVM est comparé à un PCE pour estimer la distribution de la contrainte sur un modèle CEM.

Nous avons utilisé l'implémentation de la *toolbox* « UQlab » [99] codée en Matlab. Le noyau, choisi arbitrairement, est la fonction matern-5_2. La méthode d'estimation est celle du *smooth leave-one-out*. Les paramètres C et ϵ ainsi que les paramètres du noyau sont choisis par optimisation. L'algorithme effectuant cette tâche est celui de Broyden-Fletcher-Goldfarb-Shanno (BFGS) avec 200 itérations maximum.

4. Réseau de Neurones (NN)

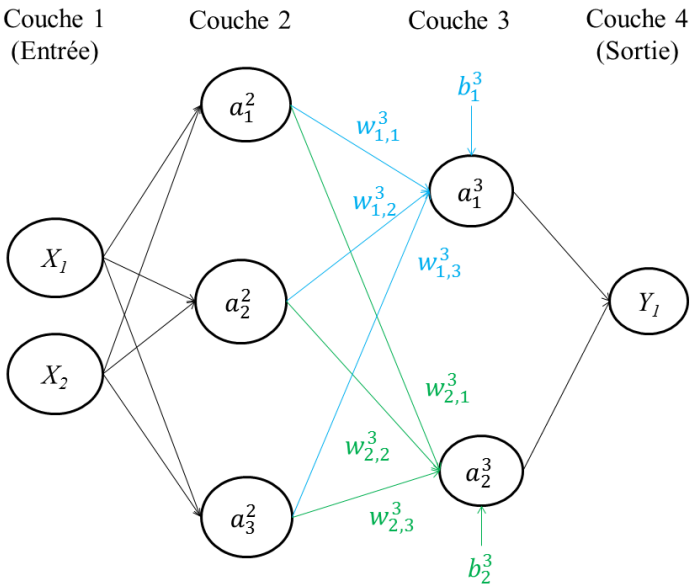

Figure 30 : Exemple d'architecture de NN avec deux variables d'entrée (X_1 et X_2), une de sortie (Y_1) et deux couches intermédiaires. Tous les poids et les biais ne sont pas représentés.

Le NN est composé de plusieurs couches de neurones : une d'entrée, une ou plusieurs intermédiaires et une de sortie. Chaque couche d'indice l contient n_l neurones d'indice $j=1,\ldots n_l$. Pour les applications de régression, le nombre de neurones dans la couche d'entrée/sortie est égal au nombre de variables d'entrée/sortie. Chaque $j^{\text{ème}}$ neurone de la $l^{\text{ème}}$ couche est caractérisé par une valeur d'activation a_j^l comprise entre 0 et 1. Les neurones des couches d'entrée prennent comme valeur d'activation la réalisation des variables d'entrée. Chaque neurone d'une couche est connecté aux neurones des couches précédentes et suivantes. La connexion entre le $j^{ème}$ neurone de la couche l et le $k^{ème}$ neurone de la couche précédente est pondérée par $w_{j,k}^l$. De plus, un biais b_j^l est également ajouté pour chaque neurone. La valeur d'activation a_j^l du $j^{ème}$ neurone de la couche l se calcule ainsi :

$$a_j^l = f\left(\sum_{k=1}^{n_{l-1}} w_{j,k}^l a_k^{l-1} + b_j^l\right), \tag{30}$$

où f est une fonction d'activation (souvent une sigmoïde [100]) afin que les a_j^l prennent une valeur nécessairement entre 0 et 1. La valeur d'activation du neurone de la couche de sortie obtenue est la valeur de la variable de sortie Y.

La Figure 30 présente un exemple de NN. L'activation du 1er neurone de la 3ème couche dépend du biais et des poids (bleu) le reliant aux neurones précédents. Il en va de même pour le 2ème neurone de la 3ème couche (vert).

Cette architecture de NN est classique et adaptée pour traiter un volume de donné modeste. Cependant, lorsque la taille d'échantillon de données d'apprentissage est très conséquente, comme dans les applications de reconnaissance d'images ou plus généralement du « Big Data », une autre architecture s'est imposée : le réseau de neurone convolutif. Bien que le principe reste similaire, l'architecture est légèrement modifiée [83]. Chaque neurone d'une couche intermédiaire n'est connecté qu'à un nombre restreint de neurones de la couche précédente. De plus, une mise en commun des poids et biais est effectuée par couche intermédiaire. Cela permet à chaque couche intermédiaire de se spécialiser dans la détection d'une caractéristique (*feature*) particulière. Le NN ainsi formé devient en quelque sorte une succession de filtres spécialisés, d'où le terme convolutif.

Le processus de construction d'un réseau de neurones commence par initialiser au hasard tous les poids et biais. Ensuite, les données d'entrée sont divisées au hasard en un ensemble dédié à l'apprentissage et un autre à la validation. L'ensemble dédié à l'apprentissage sert à la construction du NN. L'ensemble de validation permet de mesurer l'erreur du NN sur de nouvelles données et donc quantifier sa capacité prédictive. Au fur et à mesure que l'apprentissage progresse, l'erreur sur l'ensemble d'apprentissage diminue en même temps que sur celui de validation. A un stade donné, l'erreur sur l'ensemble de validation commence à augmenter alors que l'erreur continue de diminuer sur celui d'apprentissage. Les valeurs de poids/biais à cet instant, sont celles choisies. Si l'apprentissage se poursuivait, le réseau serait surentrainé, i.e. trop spécialisé sur un jeu de données particulier.

Dans le domaine des antennes, un NN a été utilisée dans [101][102][103]. Dans [104], une étude de risque d'interférence en aéronautique a été menée grâce au NN. Dans [105], le NN a permis d'optimiser le DOE. Dans [106], il s'agissait d'étudier des sources rayonnantes électromagnétiques. Dans [107], le NN a servi à étudier l'occupation des bandes de fréquences en Europe. Finalement, une intéressante association d'une méthode de réduction de dimensions (*principal component analysis*) avec le NN a été menée sur un problème de ligne de transmission [108].

Nous avons utilisé l'implémentation Matlab de la *toolbox* « Neural Net fitting ». La recherche des valeurs optimales de poids/biais a été effectuée avec l'algorithme de Levenberg-Marquardt (LM) [109] avec une régularisation bayésienne [110]. Les données sont divisées en un ensemble pour l'apprentissage (80 %), la validation (10 %). Les 10 % restants sont utilisés pour choisir le nombre de neurone par couches. En effet, si nous avons choisi arbitrairement 2 couches intermédiaires, le nombre de neurones par couche est quant à lui optimisé. La première couche peut contenir entre 10 et 40 neurones (par pas de 10) et de 5 à 35 (par pas de 10) dans la seconde couche. Toutes les combinaisons de ces deux vecteurs sont testées et celle résultant de l'erreur la plus faible est choisie. Enfin, les paramètres de l'algorithme d'optimisation LM sont les suivants : μ_0=100, μ_{inc}=5, μ_{dec}=0.5. Le maximum d'échecs de validation est fixé à 100 et le temps maximum passé est limité à 15 secondes.

5. Polynomial chaos expansion (PCE)

Introduit par Wiener [111] avec des polynômes de Hermite pour les cas où les variables d'entrée sont gaussiennes, Xiu généralise la technique dans [112]. L'idée de base des PCE [113] est que le modèle peut être approximé par une décomposition en une somme de P polynômes connus, pondérés par des coefficients inconnus $\boldsymbol{A}$. Chaque polynôme ψ est le produit de polynômes univariés d'ordre mixte (ordre max de p) pris dans une famille polynomiale appropriée. Le nombre de polynômes P (et de coefficients) augmente très rapidement avec la dimension (D).

$$SM_{PCE}(x) = \sum_{j=1}^{P} A_j \psi_j(x) \tag{31}$$

$$P = \binom{D+1}{p} \tag{32}$$

Avant de construire un PCE, la seule condition à respecter est de pouvoir construire des ensembles polynomiaux orthonormaux par rapport à la distribution des entrées. Avec une distribution ordinaire (par exemple uniforme ou normale) la littérature fournit l'ensemble à utiliser (Legendre, Hermitte (voir exemple ci-dessous)). Pour les distributions arbitraires, la procédure de Stieltjes peut être utilisée.

$$\begin{aligned} H_0(\xi) &= 1 \\ H_1(\xi) &= \xi \\ H_2(\xi) &= (\xi^2 - 1)\frac{1}{\sqrt{2}} \end{aligned} \tag{33}$$

$$H_3(\xi) = (\xi^3 - 3\xi)\frac{1}{\sqrt{6}}$$

Le processus de construction d'un PCE est le suivant :

1. Mettre en place un DOE.
2. Calculez la base polynomiale :
   - Choisir le type de polynôme orthogonal univarié (Legendre pour les entrées uniformes, Hermitte pour les entrées gaussiennes).
   - Choisir un degré maximum pour les polynômes univariés. La base est composée d'un ensemble d'un polynôme mixte ordonné. Un algorithme adaptatif peut sélectionner le degré maximal de la base polynomiale.
   - Choisir une règle de troncature (*e.g.* norme q hyperbolique). La règle définit quel polynôme univarié conserver. Chaque polynôme intermédiaire est conservé si q=1.
   - Calculer la sortie pour chaque polynôme de la base aux points d'entrée du DOE
3. Trouver le meilleur coefficient A pour chaque polynôme avec une approche par les moindres carrés aux points du DOE.

Dans [71], un PCE est utilisé dans sa version *sparse*, puis comparé en grandes dimensions à une MCS. Le PCE est aussi utilisé dans sa version *sparse* dans [114]. L'analyse de sensibilité étant facilement déduite d'un PCE, celle-ci a été réalisé dans [81] par indices de Sobol, et dans [115] avec une version bayésienne. Dans [116], un PCE est utilisé pour quantifier la réflexion électromagnétique sur des bâtiments. Enfin en CEM, le couplage entre une onde électromagnétique et une piste PCB est étudié grâce à un PCE [117].

Nous avons utilisé l'implémentation Matlab de « UQLab » [118] avec une stratégie de calcul de régression sous le moindre angle (ou LARS pour *Least Angle Regression*) et le degré maximum adaptatif (de 1 à 3). En plus de sélectionner le meilleur degré maximum, cet algorithme élimine les polynômes ayant un impact négligeable sur le résultat.

6. Krigeage (KRI)

Le KRI universel suppose que la sortie suit une réalisation de processus gaussien décrite comme la somme d'un terme de tendance et d'un terme aléatoire [119].

$$\widehat{\boldsymbol{Y}}(\boldsymbol{x}) = \boldsymbol{\beta}^T \boldsymbol{f}(\boldsymbol{x}) + \sigma^2 Z(\boldsymbol{x}), \tag{34}$$

où le terme $\boldsymbol{\beta}^T \boldsymbol{f}(\boldsymbol{x})$ représente la moyenne du processus, composée de P fonctions de base $\boldsymbol{f}$, et P coefficients de régression $\boldsymbol{\beta}$. Le terme $\sigma^2 Z(\boldsymbol{x})$ représente du bruit aléatoire, composée d'une variable aléatoire gaussienne centrée réduite $Z(\boldsymbol{x})$ multipliée par la variance du processus σ^2.

Si le processus est stationnaire (i.e. dont la moyenne est constante mais inconnue), une tendance constante est suffisante (P=1). Dans ce cas, le KRI est appelé ordinaire et la tendance s'écrit :

$$\boldsymbol{\beta}^T \boldsymbol{f}(\boldsymbol{x}) = \beta_1 f_1(\boldsymbol{x}) = \beta_1. \tag{35}$$

L'hypothèse de base du KRI est que les sorties du modèle associées à des réalisations d'entrées voisines sont plus corrélées que celles qui correspondent à des entrées plus éloignées. Les fonctions de corrélation, comme la Matérn-5/2, sont utilisées pour modéliser la relation entre la corrélation (R) et la distance euclidienne (h). Cette fonction est paramétrée par un vecteur $\boldsymbol{\theta}$ (avec autant de composants que de variables d'entrée) :

$$R(h, \boldsymbol{\theta}) = \left(1 + \frac{\sqrt{5}\,|h|}{\boldsymbol{\theta}} + \frac{5h^2}{3\boldsymbol{\theta}^2}\right) e^{-\frac{\sqrt{5}|h|}{\boldsymbol{\theta}}}, \tag{36}$$

où D est le nombre de variables d'entrée et h la distance entre deux points d'entrée $\boldsymbol{x}$ et $\boldsymbol{x}'$:

$$h = \sqrt{\sum_{i=1}^{D} \left(\frac{x_i - x_i{}'}{\theta_i} \right)^2}. \tag{37}$$

La tâche principale lors de la construction d'un KRI est de choisir le meilleur $\boldsymbol{\theta}$ selon le DOE. Grâce à un algorithme d'optimisation, comme l'algorithme génétique hybride et le gradient, le meilleur $\boldsymbol{\theta}$ peut être trouvé en minimisant la fonction de log-vraisemblance négative :

$$\boldsymbol{\theta} = arg \min_{D_{\boldsymbol{\theta}}} \left(\frac{1}{2} log(det(\boldsymbol{R})) + \frac{S_{SM}}{2} log(2\pi\sigma^2) + \frac{S_{SM}}{2}\right), \tag{38}$$

où S_{SM} est la taille du DOE. Les autres inconnues ($\boldsymbol{\beta}$ and σ^2) se déduise de $\boldsymbol{\theta}$:

$$\boldsymbol{\beta}(\boldsymbol{\theta}) = (\boldsymbol{F}^T \boldsymbol{R}^{-1} \boldsymbol{F})^{-1} \boldsymbol{F}^T \boldsymbol{R}^{-1} \boldsymbol{y}, \tag{39}$$

$$\sigma^2(\boldsymbol{\theta}) = \frac{1}{S_{SM}} (\boldsymbol{y} - \boldsymbol{F}\boldsymbol{\beta})^T \boldsymbol{R}^{-1} (\boldsymbol{y} - \boldsymbol{F}\boldsymbol{\beta}), \tag{40}$$

où la matrice $\boldsymbol{F}$ est:

$$F_{i,j} = f_j(x_i), i = 1, \dots, S_{SM}, j = 1, \dots, P. \tag{41}$$

Une fois le KRI construit, il peut être utilisé comme prédicteur à toute nouvelle entrée x_i dans l'espace de conception. La nouvelle sortie correspondante suit une distribution gaussienne de moyenne :

$$\boldsymbol{\mu}_{\hat{Y}(x_i)} = (\boldsymbol{f}(\boldsymbol{x})^T \boldsymbol{\beta} + \boldsymbol{r}(\boldsymbol{x})^T \boldsymbol{R}^{-1} (\boldsymbol{y} - \boldsymbol{F}\boldsymbol{\beta}), \tag{42}$$

et de variance :

$$\sigma^2{}_{\hat{Y}(x_i)} = \left(1 - \boldsymbol{r}^T(\boldsymbol{x}) \boldsymbol{R}^{-1}(\boldsymbol{x}) + \boldsymbol{u}^T(\boldsymbol{x}) (\boldsymbol{F}^T \boldsymbol{R}^{-1} \boldsymbol{F})^{-1} \boldsymbol{u}(\boldsymbol{x})\right) \sigma^2, \tag{43}$$

Où :

$$\boldsymbol{u}(\boldsymbol{x}) = \boldsymbol{F}^T \boldsymbol{R}^{-1} \boldsymbol{r}(\boldsymbol{x}) - \boldsymbol{f}(\boldsymbol{x}), \tag{44}$$

avec $\boldsymbol{r}$ le vecteur de corrélation croisée entre la prédiction d'un point avec tous les autres points du DOE :

$$\boldsymbol{r}(\boldsymbol{x}) = R(\boldsymbol{x}, x_i; \boldsymbol{\theta}),\ i = 1, \dots, S_{SM}. \tag{45}$$

Le prédicteur fourni par le krigeage est donc associé à une estimation de l'erreur locale (variance) de la prédiction. Le KRI a servi dans des problèmes d'optimisation [120][121], et dans de nombreux problèmes de sciences de l'ingénieur comme en mécanique [122] mais aussi en CEM [123]. Il existe de nombreuses variantes : adaptative [67], le Co-krigeage [124][125][126], Poisson-krigeage [127], krigeage bayésien [128], en association avec la méthode fiabiliste de *subset simulation* [129].

Nous avons utilisé l'implémentation du KRI Matlab de « UQlab » [119] avec tendance ordinaire, une fonction de corrélation matern-5_2, la méthode d'estimation du maximum de vraisemblance et la variante mémoire limitée de l'algorithme d'optimisation BFGS (maximum 200 itérations).

7. Le polynomial chaos expansion associé au krigeage (PCE_KRI)

Le PCE_KRI est une hybridation entre deux SMs [130] : le KRI et le PCE. Ainsi, le PCE_KRI est en fait un KRI dont la tendance est un PCE. Les avantages des deux SM (erreur locale et indices de Sobol) sont donc réunis.

Le PCE_KRI a été utilisé pour estimer des événements rare dans [131].

Nous avons utilisé l'implémentation Matlab de « UQlab » [132] avec la même caractéristique que celle utilisée pour KRI et PCE seulement.

8. Stratification Contrôlée (SC)

La SC [78] est une technique dédiée à la recherche de quantiles extrêmes. La SC échantillonne l'espace d'entrée afin de générer des événements de sortie extrêmes en beaucoup moins d'essais qu'un échantillonnage classique (Monte Carlo ou même LHS). La SC peut être vue comme une technique d'échantillonnage d'importance. La SC nécessite au préalable un modèle simple corrélé. Cependant, la question de trouver un SM approprié pour la SC est laissée en suspens dans la littérature. Dans cette thèse, nous proposons que ce modèle simple soit un SM d'interpolation stochastique. Cette sous-section est cependant dédiée à la description de la SC couplée à un SM quel qu'il soit.

Sur le budget total de n simulations, S_{SM} est réservé à la construction du SM, et les S_{CS} restants sont utilisés pour la CS. Nous avons choisi arbitrairement :

$$S_{SM} = \left\lfloor \frac{n}{2} \right\rfloor. \tag{46}$$

Nous supposons donc que le modèle simple est fourni par la construction du SM à partir des S_{SM} simulations du modèle initial. Le SM, une fois construit, est utilisé pour prédire un grand nombre de réalisations de sortie, compte tenu de son faible temps de calcul. L'échantillon de sortie prédit est décomposé en strates, d'où le nom de « stratification ». Les limites des strates sont des estimations de quantiles par un SM, d'où l'adjectif « contrôlée ». Ainsi, les limites sont les quantiles regroupés dans le vecteur $\boldsymbol{A}$:

$$\begin{cases} \boldsymbol{A} = [0, \alpha, 2\alpha, 0.5, 1], \ si\ \alpha < 0.25 \\ \boldsymbol{A} = [0, 0.5, 1 - 2(1-\alpha), \alpha, 1], \ si\ \alpha > 0.75 \end{cases}. \tag{47}$$

Le nombre de strates (n_s) est arbitraire ainsi que les limites de strates, sauf la première ou la dernière selon que l'on estime un quantile situé dans la queue inférieure ou supérieure de la distribution, respectivement. Le quantile ciblé est ici α. Pour l'ensemble des résultats présentés, nous avons utilisé 4 strates dans lesquelles un nombre spécifique de points de simulation $N_j, j = 1, \ldots, n_s$ est alloué pour un budget total de points S_{cs}.

$$N_j = \left\lfloor \frac{S_{CS}}{n_s} \right\rfloor + mod(S_{CS}, n_s) \mathbb{I}_{\{j=1 \wedge \alpha<0.25\} \vee \{j=n_s \wedge \alpha>0.75\}} \tag{48}$$

Une stratégie d'allocation plus avancée basée sur une allocation adaptative du nombre de points est également possible. Cependant, l'allocation adaptative repose sur l'estimation de l'allocation optimale, ce qui ne garantit pas une amélioration des performances, contrairement à l'allocation uniforme [21].

Le SM est donc utilisé à ce stade afin d'identifier N_j réalisations de sortie appartenant à la $j^{ème}$ strate, pour tout $j \in [1,4]$. Le modèle est appelé pour chaque entrée identifiée afin de calculer la sortie exacte associée. L'efficacité de la méthode repose alors sur le niveau élevé de corrélation entre les sorties du SM et du modèle de référence, au sens d'un opérateur de corrélation défini dans [78][21] et que nous rappelons dans (50). Si cette corrélation est suffisamment grande, la plupart des sorties du

SM associées à la $j^{ème}$ strate resteront dans la $j^{ème}$ strate du modèle exact. En fait, la performance du SM repose davantage sur la corrélation entre le SM et le modèle que sur la précision de la prédiction du SM. Autrement dit, la SC est performante si le SM approxime le modèle en termes de tendances plus qu'en valeurs absolues. Dans de telles conditions, une entrée produisant un événement extrême avec le SM est susceptible de produire également un événement extrême du modèle réel. La convergence vers l'estimation de quantiles extrêmes peut alors être considérablement accélérée.

Enfin, la *cdf* estimée de la contrainte par la méthode de CS peut être calculée :

$$F_Y(y) = \sum_{j=1}^{n_s} \left((A_{j+1} - A_j) \frac{1}{N_j} \sum_{i=1}^{N_j} 1_{Y_i^{(j)} \leq y} \right). \tag{49}$$

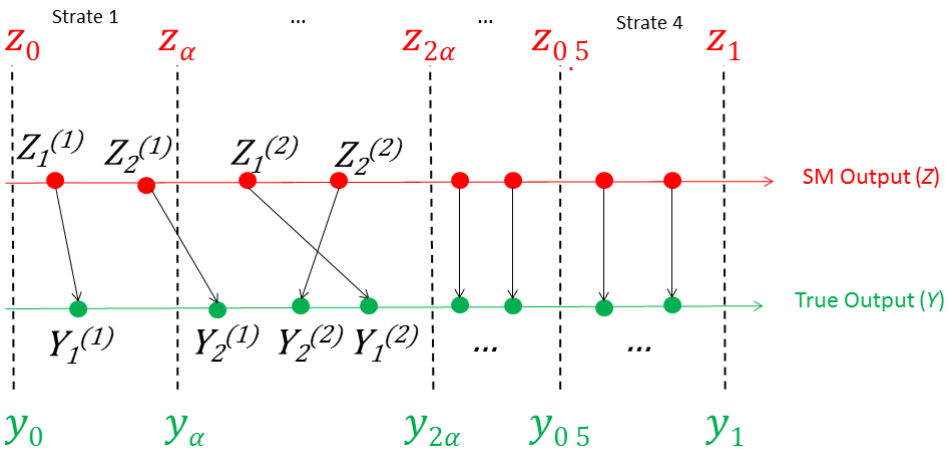

Figure 31 : Schéma de principe de la stratification contrôlée dans le cas où n_s=4, N_j=2 $\forall\, j$ et $\boldsymbol{A} = [0, \alpha, 2\alpha, 0.5, 1]$.

La Figure 31 illustre le principe de la stratification où l'on représente la sortie du SM, appelée Z (axe rouge). Ces sorties sont stratifiées par les quantiles de Z selon des probabilités regroupées dans le vecteur **A**. Dans cette figure, nous représentons arbitrairement 2 réalisations par strates. Pour chaque réalisation d'entrée qui correspond à la sortie Z du SM, la sortie du vrai modèle est calculée (variable *Y*). Certaines réalisations de Y (axe vert) restent dans la même strate. C'est le cas pour celles issues des strates 2, 3 et 4 ainsi que pour la première réalisation de la première strate i.e. $Y_1^{(1)}$. Cependant, ce n'est pas toujours le cas puisque d'autres réalisations peuvent entraîner un changement de strate. Par exemple, le modèle appelé pour la deuxième réalisation de la première strate i.e. $Z_2^{(1)}$ aboutit à $Y_1^{(1)}$ qui appartient en fait à la deuxième strate. Plus le nombre de réalisations qui changent de strates est faible, plus la SC sera performante. Il existe un moyen de quantifier précisément ce phénomène grâce à un indicateur appelé corrélation. La corrélation ρ, évaluée pour un quantile ciblé de probabilité α, se calcule ainsi :

$$\rho(\alpha) = \frac{\mathbb{P}\{Y \leq y_\alpha | Z \leq z_\alpha\}\alpha - \alpha^2}{\alpha - \alpha^2} 100. \tag{50}$$

Si *Y* et *Z* sont indépendants, la corrélation est nulle. En effet :

$$\mathbb{P}\{Y \leq y_\alpha \,|Z \leq z_\alpha\} = \mathbb{P}\{Y \leq y_\alpha\} = \alpha$$
$$\Rightarrow \rho = 0\,.$$

Si aucune valeur extrême de *Z* ne correspond à une valeur extrême de *Y*, alors la corrélation est négative. En effet :

$$\mathbb{P}\{Y \leq y_\alpha \,|Z \leq z_\alpha\} = 0 \tag{51}$$

$$\Rightarrow \rho = \frac{-\alpha^2}{\alpha-\alpha^2} 100 < 0 .$$

Enfin, si chaque valeur extrême de Z correspond à une valeur extrême de Y, alors la corrélation est parfaite. En effet :

$$\begin{gathered} \mathbb{P}\{Y \leq y_\alpha \,|Z \leq z_\alpha\} = 1 \\ \Rightarrow \rho = \frac{\alpha-\alpha^2}{\alpha-\alpha^2} 100 = 1. \end{gathered} \tag{52}$$

III. Modèles utilisés

Dans cette section nous allons présenter deux modèles analytiques qui seront très souvent utilisés dans ce chapitre. Il s'agit de modèles où la sortie est calculée analytiquement, connaissant l'entrée. Nous aurons besoin dans ce chapitre de tester et valider des méthodes statistiques, ce qui nécessite une référence. Celle-ci pourra être facilement et rapidement obtenue par un de ces modèles analytiques. Dans un objectif de validation de méthode, obtenir des valeurs exactes de contrainte n'est pas important. Ainsi, l'exactitude de ces modèles analytique n'a pas besoin d'être évaluée. Lorsque l'objectif sera de mettre en pratique les méthodes sur des problèmes plus concrets, nous utiliserons un modèle *full-wave*, coûteux en temps de calcul mais considéré comme exact.

1. Modèle PCB

Le premier modèle utilisé calcule l'amplitude du champ électrique lointain, rayonné par une piste PCB. Celle-ci est chargée à chaque extrémité, au-dessus d'un plan de masse infini comme indiqué dans la Figure 32. Les détails du calcul de la sortie du modèle à partir des paramètres d'entrée se trouvent en Annexes.

Ce modèle est défini par 11 variables aléatoires d'entrée, présentées dans le Tableau 13. Chaque entrée suit une distribution gaussienne centrée sur sa valeur nominale avec un écart-type relatif de 10 %. Les valeurs nominales (i.e. la moyenne d'une variable) sont choisies de manière à ce que la piste apparaisse comme une ligne de transmission en quart d'onde désadaptée.

Un échantillon de référence a été calculé à partir de 10^6 réalisations d'entrée permettant de calculer les quantiles visés. Un histogramme des sorties de l'échantillon de référence est représenté sur la Figure 33 ainsi que les trois quantiles de cet échantillon : $q_{90\%}$, $q_{99\%}$, $q_{99,9\%}$. La précision estimée de ces quantiles de référence est respectivement ±0.0093 %, ±0.019 % et ±0.044 %. Cette précision est estimée à partir de la distribution empirique avec un intervalle de confiance à 95% grâce à la fonction Matlab *ecdf*. Notons que ce niveau de précision est bien supérieur à ce qui sera obtenu par les méthodes statistiques utilisées par la suite. Nous pouvons donc bien considérer ces quantiles de référence comme exacts.

Nous représentons dans la Figure 34 la sortie du modèle lorsque toutes les variables sont fixées à leur valeur nominale à l'exception de la fréquence (courbe noire). Nous constatons un rayonnement maximal à 404 *MHz*, ce qui correspond à la ligne de transmission en quart d'onde, comme attendu. Lorsque toutes les variables sont aléatoires nous obtenons un nuage de points. Les points colorés en bleu correspondent à des valeurs de champ électrique situées en dessous-du quantile à 90 % (69.95 *dBμV/m*). Lorsque les valeurs du champ sont supérieures à ce seuil, les points correspondants sont rouges.

Tableau 13 : Entrées du modèle PCB.

nom	Description	Nominale μ	Distribution
f	fréquence	$404\ MHz$	
h	Hauteur de la piste	$0.775\ mm$	
W	Largeur piste	$0.51\ cm$	
l	Longueur piste	$10.16\ cm$	
ε_r	permittivité du substrat	4.6	
V_s	source de tension	$1\ V$	$\mathcal{N}(\mu, \sigma = 0.1\mu)$
Z_s	Impédance source	$50\ \Omega$	
Z_l	Impédance de charge	$1\ \Omega$	
r	Distance où le champ est mesuré	$3\ m$	
θ	Angle élevation	2π	
φ	Angle azimutal	2π	

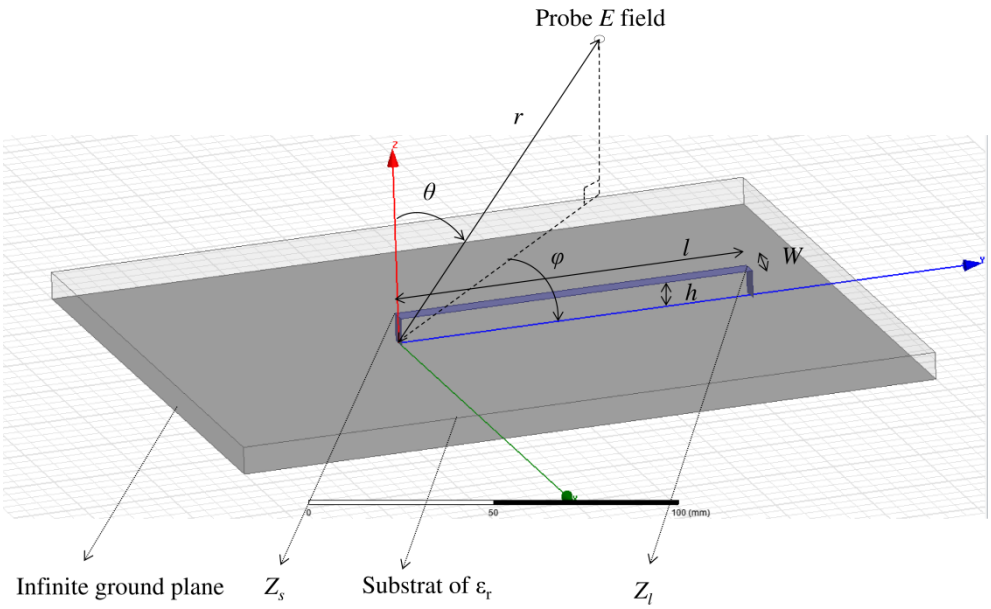

Figure 32 : Schéma de la piste PCB (extrait de [133]).

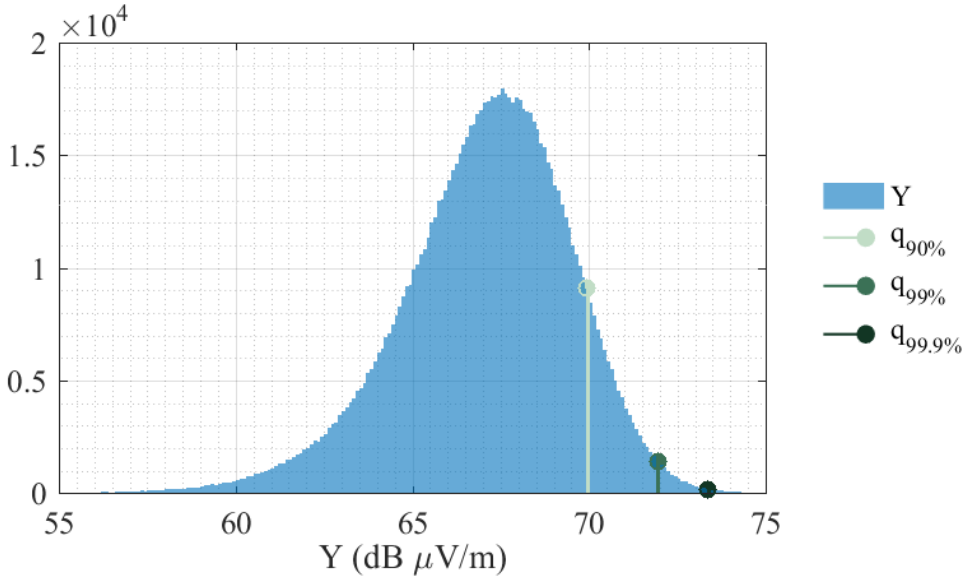

Figure 33 : Modèle de rayonnement PCB : histogramme des 10^6 réponses du modèle vrai du rayonnement des PCB et des quantiles réels qui seront estimés.

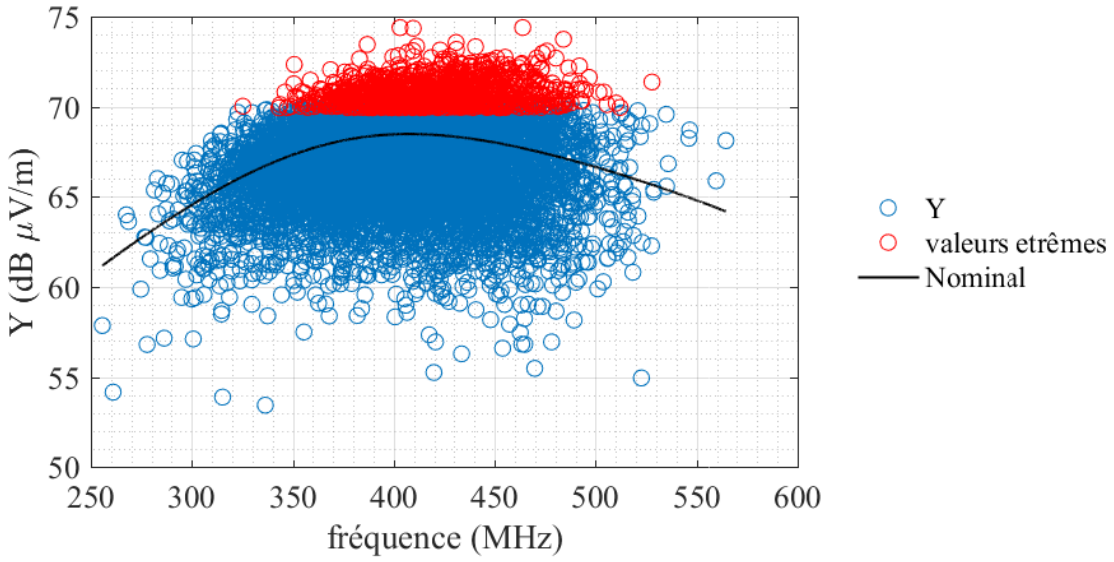

Figure 34 : Modèle de rayonnement PCB : champ électrique rayonné pour 10000 entrées aléatoires (points rouges pour $Y>y_{90\%}$) et variation du champ électrique rayonné en fonction de la fréquence tandis que les autres entrées sont fixées à leur valeur nominale.

2. Modèle S_{11}

Le deuxième modèle étudié est le coefficient de réflexion (S_{11} en linéaire) d'un circuit RLC série représenté dans la Figure 35.

La sortie de ce modèle est calculée ainsi :

$$Y = \left| \frac{R + j2\pi f \left(L - \frac{1}{C(2\pi f)^2} \right) - 50}{R + j2\pi f \left(L - \frac{1}{C(2\pi f)^2} \right) + 50} \right| \tag{53}$$

Les 4 entrées aléatoires sont décrites dans le Tableau 14. L'impédance de charge est fixe

(*Zl=50* Ω) ainsi que la tension d'entrée (1 *V*). Les valeurs nominales sont choisies pour atteindre une résonance à 500 MHz. Une résonance parfaite n'est pas réaliste. Cependant, cela permet de créer un modèle difficile à approximer pour un SM à cause de la forte non-linéarité induite par le phénomène de résonance.

La distribution des sorties du modèle est représentée sur la Figure 36 à partir d'un échantillon de référence d'1 million de réalisations de sortie. Pour $q_{10\%}$ et $q_{1\%}$, la précision est de ±0.55 % et ±1.1 %, respectivement. A mesure que la probabilité diminue, il y a de moins en moins de réalisations parmi l'échantillon de référence qui sont inférieures au quantile. C'est pourquoi la précision est moins bonne pour les quantiles plus extrêmes. Ces précisions sont encore bien meilleures que celles qui seront obtenues par les méthodes traitées dans ce document. Par conséquent, ces quantiles de référence sont considérés comme exacts.

Nous représentons dans la Figure 37, le coefficient de réflexion (donc la sortie du modèle) lorsque toutes les variables sont fixées à leur valeur nominale à l'exception de la fréquence (courbe noire). Lorsque toutes les variables sont aléatoires nous obtenons un nuage de points (rouges pour ceux au-dessous du 10 % quantile, bleus sinon).

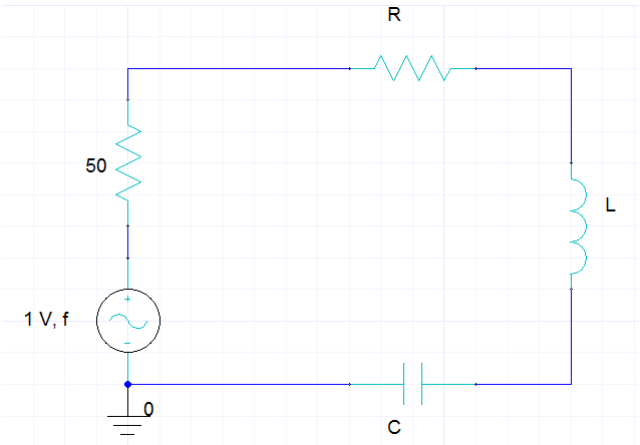

Figure 35 : Modèle de circuit RLC série.

Tableau 14 : Entrées du modèle S11.

Nom	Description	Nominale μ	Distribution
f	fréquence	$500\ MHz$	$U(100{,}900)\ MHz$
R	Résistance	$50\ \Omega$	$U(\mu - 0.1\mu, \mu + 0.1\mu)$
L	Inductance	$1.5\ pF$	
C	Capacité	$67.5\ nH$	

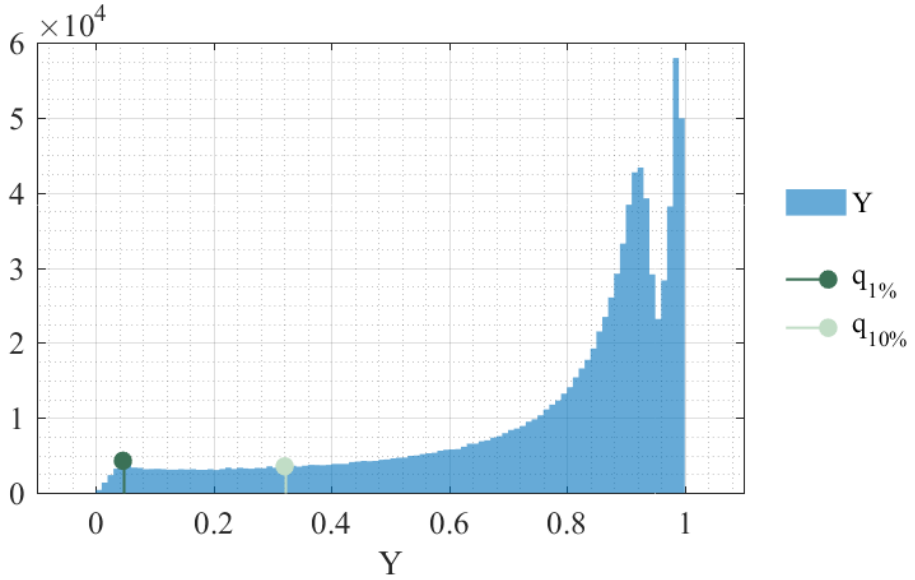

Figure 36 : Modèle S_{11} : histogramme des 10^6 réponses du modèle vrai et des quantiles vrais qui seront estimés.

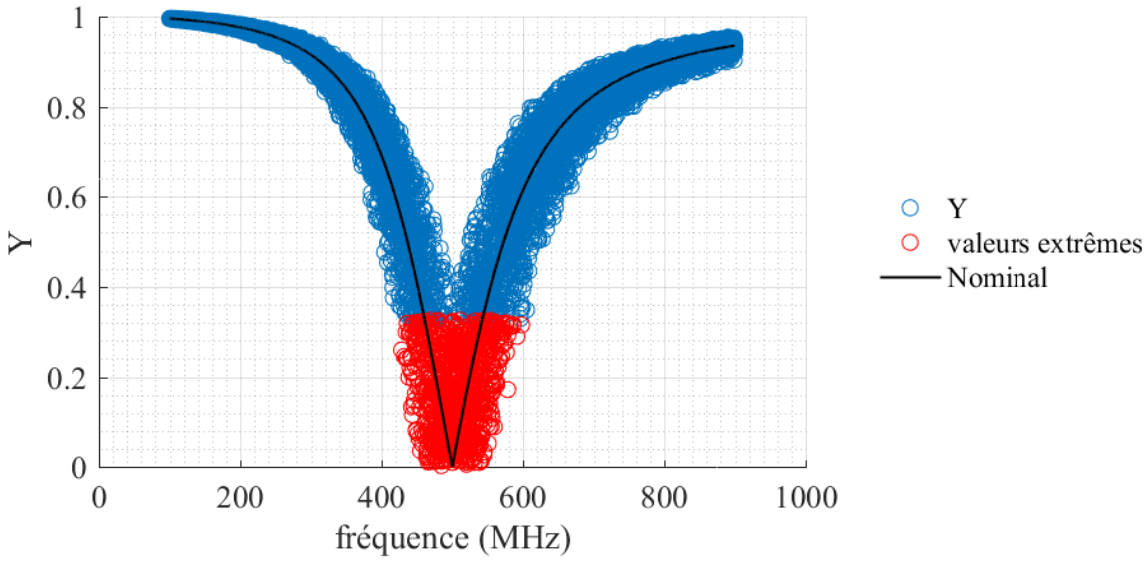

Figure 37 : Modèle S_{11} : 10000 réalisations de sortie (en linéaire). Les points rouges représentent des sorties considérées comme extrêmes i.e. inférieures à un quantile (ici 10 %). La courbe noire correspond au cas où toutes les entrées sont fixées à leur valeur nominale à l'exception de la fréquence.

IV. Comparaison des performances des métamodèles

Nous allons à présent comparer les cinq SMs présentés dans la section II, c'est-à-dire le *Support Vector Machine* (SVM), le *Réseau de Neurone* (NN), le *Polynomial Chaos Expansion* (PCE), le *krigeage* (KRI), et le krigeage avec tendance PCE (PCE_KRI). Les modèles déterministes et supposés exacts que les SMs sont censés approximer sont les deux modèles présentés dans la section III, c'est-à-dire le modèle calculant le coefficient de réflexion S_{11} d'un circuit RLC et celui calculant le champ électrique rayonné par une ligne microruban d'un PCB. L'objectif est de sélectionner le SM le plus approprié pour être utilisé par la SC.

1. Critères de performance

Les performances de chaque SM sont évaluées selon deux critères :

- L'erreur relative spécifique (*Err*) entre le modèle et le SM, spécifique car cantonnée aux valeurs extrêmes.
- La corrélation (ρ) entre le modèle et le SM, au sens de la stratification contrôlée.

Le critère de corrélation est le plus important. Tant que la corrélation entre le SM et le modèle est suffisamment élevée, la SC peut être très efficace pour estimer les valeurs extrêmes. En fait, l'erreur du SM n'affecte pas systématiquement les performances de la SC. Si l'erreur est suffisamment faible, un SM peut être utilisé directement pour estimer les quantiles extrêmes (propagation de l'incertitude dans la SM). Il peut ne pas être intéressant d'utiliser le CS en plus du SM dans ce cas. Au contraire, si l'erreur est importante mais que la corrélation reste bonne, le SC peut être meilleur que le SM seul.

Un SM est construit avec un petit échantillon de taille *n*. A partir du SM constitué, un second échantillon de test, issu de tirages indépendants et de taille *N* beaucoup plus importante (N=10000>>n) est ensuite utilisé pour évaluer les performances des SMs. Parmi ces *N* réalisations, *Next* sont extrêmes, i.e. appartenant à la queue de distribution. Ces réalisations extrêmes sont celles qui sont supérieures au quantile y_α (estimé empiriquement à partir des *N* réalisations du SM) lorsque α est proche de 1 ou inférieur à ce quantile si α est proche de 1. L'erreur spécifique associée aux valeurs extrêmes se calcule ainsi :

$$Err = \sqrt{\frac{1}{N_{ext}} \sum_{i=1}^{N_{ext}} \frac{\left(y_{ext_i} - \hat{y}_{ext_i}\right)^2}{y_{ext_i}}}. \tag{54}$$

La corrélation entre la variable de sortie du modèle vraie (*Y*) et celle du SM (*Z*) au sens de la SC est calculée avec l'ensemble des *N* réalisations du SM à des probabilités données α. Si y_α et z_α sont les quantiles α du modèle et du SM, respectivement, la corrélation est donnée par (50). On rappelle qu'idéalement on souhaite obtenir $\rho(\alpha)$ le plus proche possible de 1.

Ces indicateurs de performances, *Err* et $\rho(\alpha)$ dépendent de l'échantillon initial. Par conséquent, les distributions d'*Err* et de $\rho(\alpha)$ sont estimées à partir de 100 DOEs indépendants de taille *n*.

2. Modèle PCB : Résultats

La Figure 38 représente la distribution de l'erreur calculée au sens de (14) pour différentes tailles de DOE (n=50,100,200), et ce pour chaque type de SM. Le quantile recherché est α=90 %. La distribution des erreurs est représentée sous forme de boîtes à moustaches. Les rectangles de ces boîtes à moustaches délimitent les quantiles à 25 % et 75 %, et le trait horizontal scindant le rectangle en deux représente la médiane. Enfin les traits horizontaux extérieurs aux rectangles représentent les quantiles à 10 % et 90 %. Nous remarquons tout d'abord que les erreurs relatives sont déjà faibles, même pour *n* faible. Au fur et à mesure que *n* augmente, celles-ci diminuent et sont moins dispersées. Le PCE_KRI possède l'erreur la plus faible, suivi du KRI. Le SVM est plus performant que le PCE et le NN pour les petits *n* mais devient le moins performant de tous les SMs pour de plus grandes valeurs de *n* (100 et 200). Par conséquent, le classement de ces différents SMs selon leur erreur (de la plus faible à la plus grande) est : PCE_KRI, KRI, SVM, NN, PCE pour les deux plus faibles valeurs de *n* (50 et 100) et pour la plus grande valeur de *n* (200) SVM tombe à la dernière place. Nous pouvons cependant conclure qu'aucune méthode ne se détache significativement d'une autre selon ce critère, pour ce modèle et pour le quantile recherché (α=90 %).

La corrélation est représentée sur la Figure 39 pour le petit échantillon (n=50), en fonction du quantile recherché α fixé successivement à 91 %, 95 % et 99 %. Au fur et à mesure que α augmente, le

niveau de corrélation diminue quel que soit le SM choisi. Par ailleurs son estimation est plus dispersée. C'est un résultat attendu compte tenu de la faible taille de l'échantillon. Le niveau de corrélation observé est le plus élevé pour les techniques PCE_KRI et KRI. La méthode NN aboutit à une corrélation du même ordre ou légèrement inférieure. Le PCE possède la corrélation la plus faible et la SVM possède une performance à peine plus élevée. Le classement de la corrélation la plus élevée à la plus faible est donc assez similaire au classement de l'erreur. Cependant, les niveaux de corrélation obtenus par PCE_KRI, KRI et NN sont significativement supérieurs du point de vue de l'impact potentiel sur la méthode SC.

Bien que le classement selon l'erreur soit le même que le classement selon la corrélation, l'erreur est déjà très faible (≤ 2 %) alors que la corrélation n'est pas très élevée. Nous soupçonnons que dans le cas du PCB, la SC ne sera pas nécessaire car le PCE_KRI approxime déjà bien le modèle pour les valeurs extrêmes. En fait, en raison d'une corrélation modérée, la SC avec le PCE_KRI pourrait être moins performante que le PCE_KRI seul.

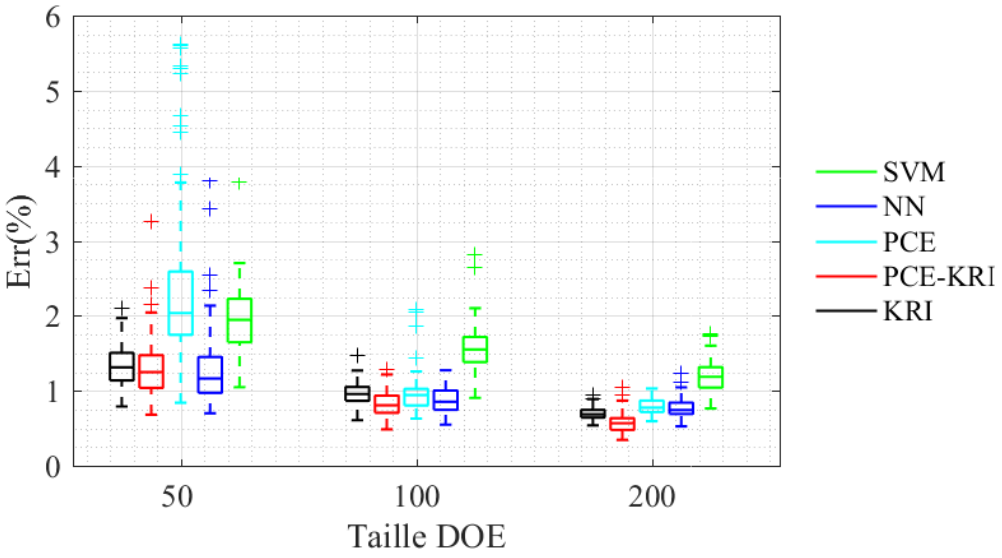

Figure 38 : Modèle PCB : distributions d'erreurs généralisées pour les valeurs extrêmes ($Y>y_{90\%}$) en fonction des tailles de DOE (n).

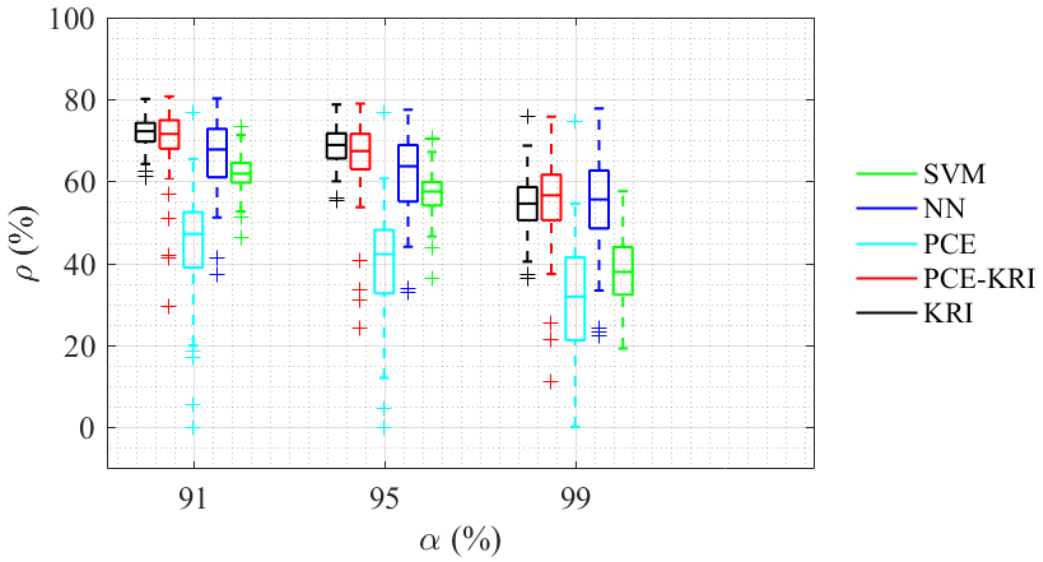

Figure 39 : Modèle PCB : distribution de la corrélation pour un DOE de taille n=50 en fonction de la probabilité α.

3. Modèle S11 : Résultats

Dans la Figure 40, les erreurs relatives de chaque SM sont très élevées. Le quantile recherché est cette fois α=10 %. Ce sont les méthodes SVM et PCE qui présentent des niveaux d'erreurs les plus importants. KRI, PCE_KRI et NN sont similaires bien que les performances de NN soient légèrement meilleures. Le classement selon l'erreur du SM le plus performant au moins performant est donc : (PCE_KRI, KRI, NN), PCE, SVM. KRI est presque identique à PCE_KRI, qui est un KRI avec une tendance PCE. Ajouter un modèle de tendance au KRI par un PCE n'améliore pas les performances par rapport au KRI qui considère une tendance constante. Soit le processus aléatoire est stationnaire (tendance constante) et le PCE ne l'estime pas mieux que le KRI seul, soit il y a une tendance mais le PCE estime qu'elle est constante.

Dans la Figure 41, le classement de la corrélation est le même que pour l'erreur. Le quantile recherché α est fixé successivement à 1 %, 5 % et 10 %. De nouveau, les SMs de type PCE_KRI et NN se distinguent par leur performance. Les corrélations négatives rencontrées avec la technique SVM, ne sont pas impossibles, mais évidemment pas souhaitées.

Les erreurs moyennes aux valeurs extrêmes sont extrêmement élevées tandis que la corrélation moyenne atteint des niveaux relativement élevés. Dans le cas du modèle S11, nous soupçonnons que le SC pourrait apporter un gain de performance par rapport à un SM seul.

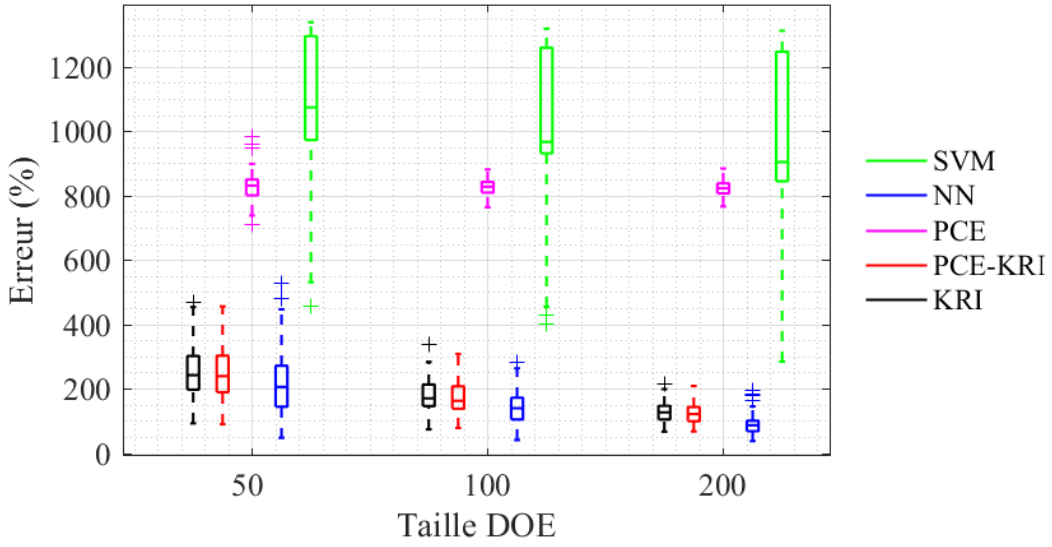

Figure 40 : Modèle S11 (RLC réflexion): distributions d'erreurs généralisées pour les valeurs extrêmes ($Y \leq y_{10\%}$) en fonction des tailles de DOE (n).

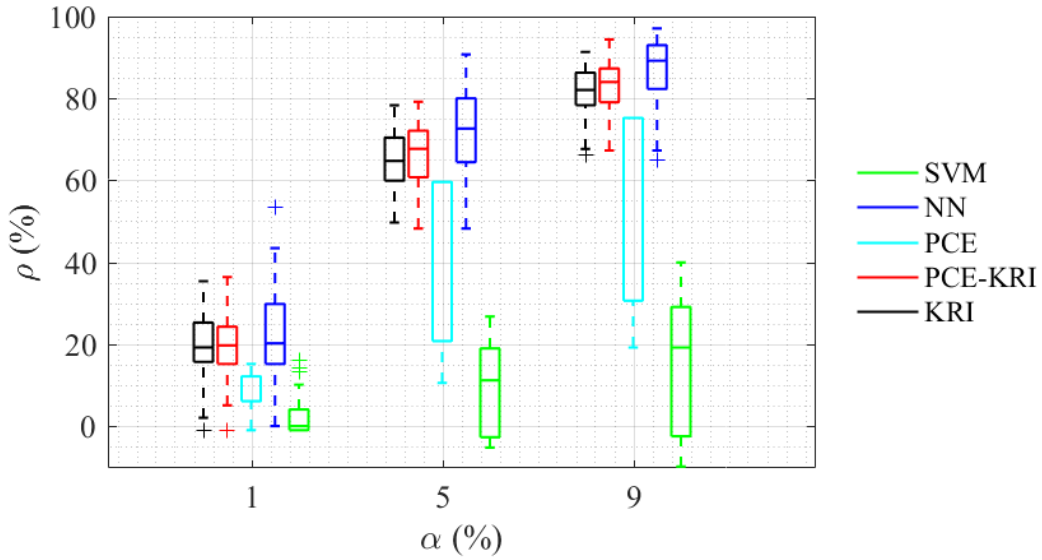

Figure 41 : Modèle RLC : distribution de la corrélation pour un DOE de taille n=50 en fonction de la probabilité α (1 %, 5 % et 9 %).

4. Sélection d'un métamodèle : Le krigeage

Nous avons donc comparé les performances de cinq SMs, candidats à l'approche SC, afin d'estimer des quantiles extrêmes, sur deux modèles différents et pour des DOEs de différentes tailles. Cette comparaison est basée sur deux critères, tous deux associés à l'analyse des valeurs extrêmes du SM du point de vue de l'erreur relative et de la corrélation ou tendance. Le classement de ces cinq SMs est globalement similaire pour ces deux critères.

Dans la plupart des cas, le PCE_KRI a obtenu les meilleurs résultats, suivis de près ou égalé par le KRI. Tout comme le KRI, le PCE_KRI fournit une estimation de l'erreur locale de sa prédiction. Une amélioration adaptative du DOE pourrait être facilement mise en œuvre et améliorerait donc très probablement les performances. De plus, le PCE_KRI possède également des indices de sensibilité, contrairement au KRI. Dans le cas de dimensions très élevées, on peut utiliser la variante à faible

densité de PCE [134]. Même si le NN est attractif, il ne fournit pas d'erreur locale ni d'analyse de sensibilité.

Nous avons découvert dans l'exemple d'application du circuit RLC qu'avec le meilleur SM, i.e. le PCE_KRI, l'erreur relative spécifique de prédiction du SM dans la queue de distribution de sortie est très importante (supérieure à 120 %) mais que la corrélation reste cependant élevée (supérieure à 58 % pour un petit échantillon de 50 réalisations). Dans ce contexte où la réponse extrême du système étudié possède un comportement relativement non linéaire, l'utilisation de la SC pourrait pallier les insuffisances du SM. Dans l'exemple du rayonnement du PCB, l'erreur de prédiction de valeurs extrêmes par l'ensemble des SMs est seulement résiduelle (≤ 2 % avec le meilleur SM i.e. le PCE_KRI) tandis que la corrélation n'est pas très élevée. Pour ce qui concerne ce cas d'étude, il semble donc a priori que la SC n'apporte pas de plus-value à l'utilisation du SM seul.

Au vu de ces résultats et des arguments précédents, nous pourrions être logiquement tentés de choisir le PCE_KRI. Cependant, le temps de calcul de la partie PCE retarde considérablement la construction du PCE_KRI par rapport au KRI. C'est pourquoi nous retenons dans la suite de ce chapitre le KRI comme support de la stratification contrôlée. Le KRI peut également être encore amélioré en échantillonnant de manière adaptative l'espace d'entrée [135], mais nous nous contentons de sa version non-adaptative. Le temps consacré à la construction du KRI adaptatif peut devenir prohibitif pour de très grands échantillons et/ou des problèmes de grandes dimensions. Dans ce cas, il faudra utiliser des techniques de réduction d'ordre par une analyse a priori de sensibilité ou des variantes spécifiques comme le *Partial Least Square kriging* [136].

V. Intérêt de la Stratification Contrôlée par rapport au krigeage seul

La stratification contrôlée accélère la convergence de l'estimation empirique standard vers les valeurs extrêmes, réduisant ainsi le nombre d'appels du modèle coûteux en temps de calcul. Cependant, la stratification contrôlée nécessite un modèle simple (c'est-à-dire dont le temps de calcul est rapide) avec une corrélation suffisante avec le modèle initial. Le KRI a été identifié dans la section précédente comme le meilleur compromis, compte tenu de ses performances et de la rapidité de sa construction. Nous allons à présent étudier la valeur ajoutée de la stratification contrôlée combinée au KRI par rapport à l'utilisation du seul KRI en vue de déterminer un quantile extrême.

1. Algorithme de comparaison et critères de performance

Pour les deux modèles (S_{11} et PCB), l'algorithme conçu pour la comparaison suit quatre étapes. Il est exécuté plusieurs fois afin d'évaluer une statistique d'erreur d'estimation du quantile. En effet l'erreur d'estimation dépend de l'échantillon aléatoire initial fourni par le DOE. Ces quatre étapes sont les suivantes :

- Obtenir un DOE de taille n : échantillonnage des entrées, puis calcul des sorties associées par le modèle exact.
- Estimer le quantile ciblé à l'aide d'une des trois méthodes
  - EE : estimation empirique direct.
  - KRI : estimation indirecte à partir des sorties (en grand nombre) du SM de type KRI constitué à partir de l'échantillon d'apprentissage obtenu.
  - K-SC : estimation à l'aide de la méthode SC, basée sur l'exploitation du SM de type KRI.
- Calculer l'erreur relative entre le quantile estimé et le quantile réel :

$$Eq = \frac{\hat{q}_\alpha - q_\alpha}{q_\varepsilon} \cdot 100 \tag{55}$$

- Calculer les statistiques d'Eq pour les trois méthodes. Les statistiques choisies pour comparer les erreurs (Eq) sont l'erreur moyenne :

$$\theta_1 = \hat{\mu}_{Eq}, \tag{56}$$

et l'erreur absolue maximale :

$$\theta_2 = \hat{\mu}_{|Eq|} + 1.96\,\hat{\sigma}_{|Eq|}\,. \tag{57}$$

2. Modèle S11 : Méthodologie et résultats

Pour le modèle S11, nous répétons 32 fois les 4 phases de calcul présentées au début de la section V pour établir les statistiques θ_1 et θ_2. Chaque simulation Monte Carlo prend du temps, surtout pour les grandes tailles de DOE. Par exemple, nous indiquons dans le Tableau 15 le temps consacré à la construction d'un seul KRI, qui lorsque la taille de DOE augmente n'est plus négligeable. En conséquence, nous avons dû limiter à 32 le nombre de tirages de Monte Carlo, ce qui correspond à la génération de 32 réalisations d'erreurs. L'inférence de θ_1 et θ_2 a ensuite été réalisée à l'aide d'une méthode de Bootstrapping [54] et d'un ajustement d'une fonction noyau (*kernel*) [137] grâce aux fonctions Matlab *bootstrap* et *ksdensity*.

La technique de Bootstrapping est une méthode de ré-échantillonnage avec remplacement pour estimer la distribution d'une statistique. Nous avons ré-échantillonné 1000 sous-échantillons à partir des 32 échantillons originaux d'erreurs, calculé les statistiques pour chaque nouvel échantillon afin d'obtenir 1000 réalisations de θ_1 et θ_2. L'histogramme de ces réalisations est ensuite ajusté à une densité de noyau gaussien (*gaussian kernel fit*). Enfin, un intervalle de confiance à 95 %, ainsi que la réalisation la plus probable (maximum de la *pdf* du noyau) peuvent être déduite de la distribution ajustée.

Nous traçons l'estimateur de la moyenne des erreurs θ_1 et l'estimateur de l'erreur maximale θ_2 pour différentes tailles de DOE lorsque le quantile cible est $q_{10\%}$ (Figure 42 pour θ_1 et Figure 43 pour θ_2) et $q_{1\%}$ (Figure 44 pour θ_1 et Figure 45 pour θ_2). Dans ces figures, le symbole (carré pour θ_1 ou étoile pour θ_2) correspond à la réalisation la plus probable de l'estimateur tandis que la barre d'erreur correspond à l'intervalle de confiance à 95 %, i.e. les quantiles à 2.5 % et 97.5 % de la distribution de l'estimateur. Dans les quatre figures, lorsque l'on augmente la taille des DOE, les erreurs tendent à diminuer.

Plus précisément, concernant l'estimation du quantile $q_{10\%}$ (Figure 42 et Figure 43) on peut distinguer deux régions de n :

- n=[50,150] : aucune méthode ne fonctionne très bien (risque de valeurs élevées sur θ_2). Nous notons également que θ_1 est élevé pour le KRI, ce qui signifie que l'estimation de $q_{10\%}$ est biaisée positivement pour n=50.
- n=[250,350,450] : l'erreur maximale θ_2 obtenue par le KRI est bien inférieure à *celle* obtenue en recourant à la SC (K-SC). Par conséquent, le KRI est plus performant que le K-SC, qui reste cependant plus performant que l'EE.

Concernant l'estimation du quantile $q_{1\%}$ (Figure 44 et Figure 45) on peut aussi distinguer deux régions de n :

- n=1000 : aucune des trois méthodes ne donne de bons résultats. De nouveau la valeur de θ_1 obtenue pour le KRI montre que l'estimation de $q_{1\%}$ est très fortement biaisée positivement.
- n=[2000,3000,4000] : l'erreur maximale θ_2 obtenue par la méthode du KRI est similaire à celle qui résulte de l'EE. Par contre cette erreur maximale est cette fois bien plus faible à l'aide de l'hybridation KSC. De plus, la moyenne θ_1 obtenue par KRI est de nouveau synonyme d'un fort biais d'estimation. Par conséquent, K-SC surpasse les performances des méthodes EE et KRI.

Le modèle est fortement non linéaire en raison du phénomène de résonance. La taille du DOE nécessaire pour atteindre une précision raisonnable est donc très grande. Par conséquent, le temps consacré à la construction d'un KRI n'est pas négligeable, comme le montre le Tableau 15. Puisque

l'hybridation K-SC n'a besoin que de la moitié de la taille du DOE nécessaire pour le KRI, le temps de calcul associé à l'élaboration de K-SC est beaucoup moins long que lorsque l'on utilise le KRI seul.

On peut donc conclure que pour le modèle de S_{11}, la SC apporte un gain de performance important par rapport au KRI seul pour estimer un quantile extrême.

Tableau 15 : Temps de simulation consacré à la construction d'un seul KRI pour un seul DOE.

Taille de DOE	100	200	400	800	1600	3200	6400
Temps (minutes)	0.42	0.47	0.65	1.28	3.54	12.81	106.3

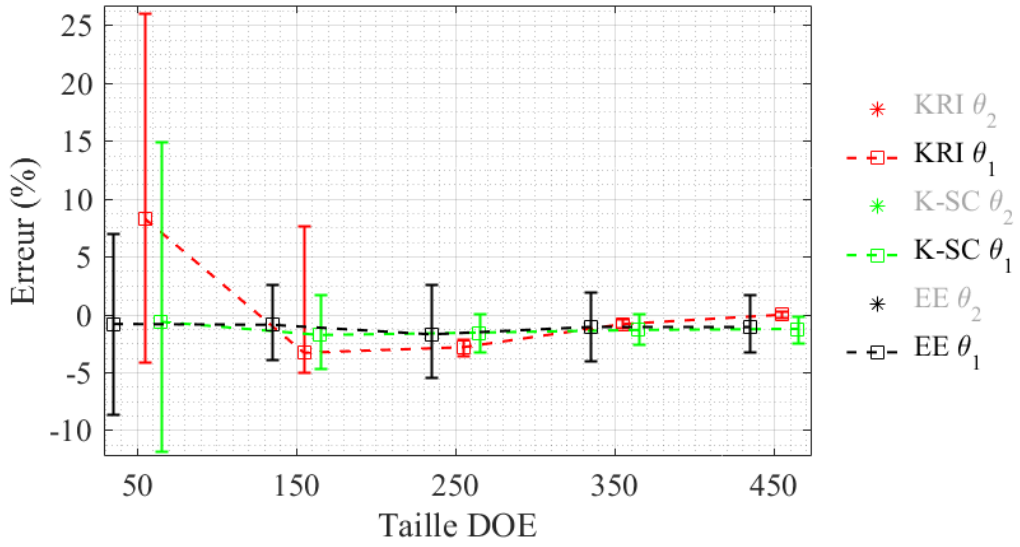

Figure 42 : Modèle S11, estimation de $q_{10\%}$. Erreur relative moyenne (θ_1) par estimation empirique (EE), krigeage (KRI) et krigeage + Stratification Contrôlée (K-SC). Les résultats pour θ_2 se trouvent dans la Figure 43.

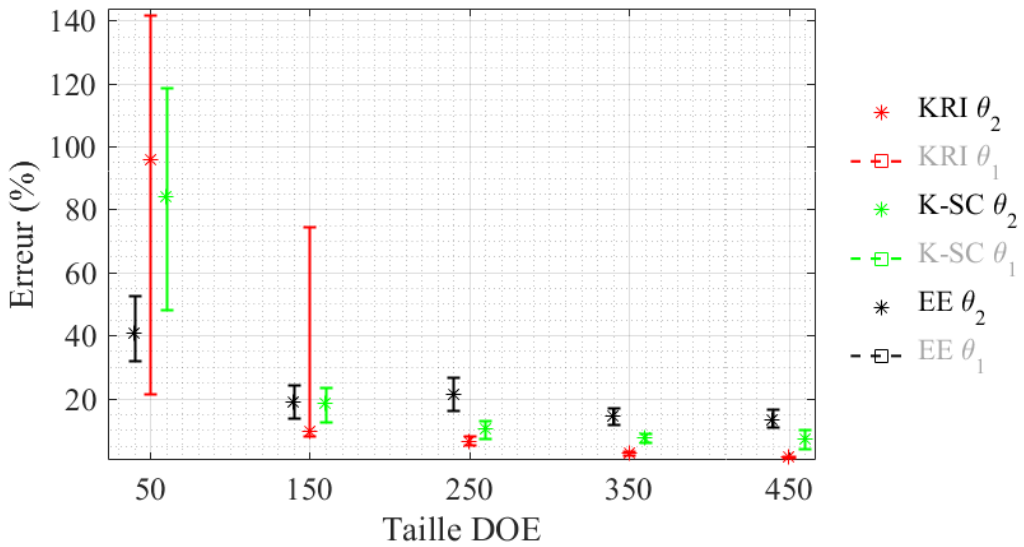

Figure 43 : Modèle S11, estimation de $q_{10\%}$. Erreur relative absolue maximale (θ_2) par estimation empirique (EE), krigeage (KRI) et krigeage + Stratification Contrôlée (K-SC). Les résultats pour θ_1 se trouvent dans la Figure 42.

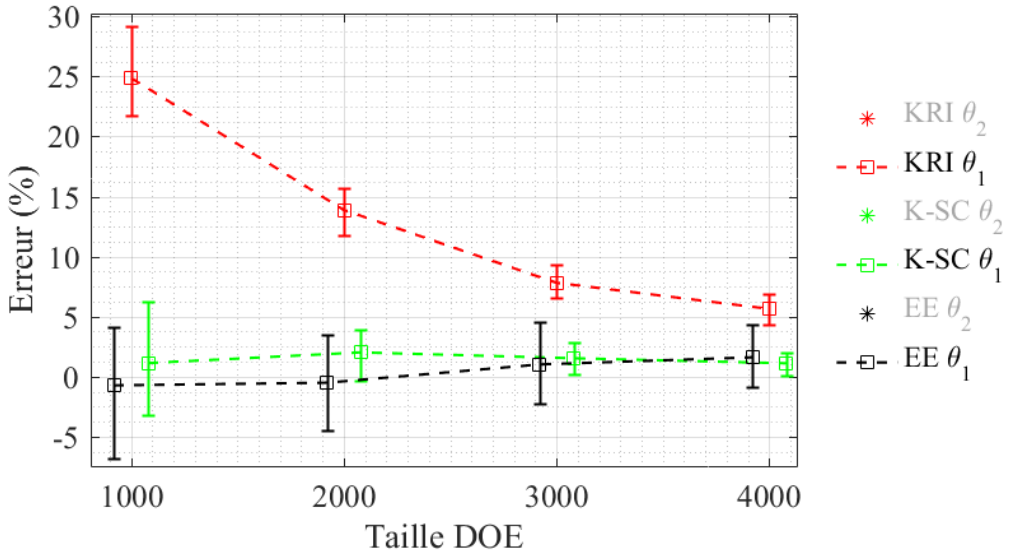

Figure 44 : Modèle S11, estimation de $q_{1\%}$. Erreur relative moyenne (θ_1) par estimation empirique (EE), krigeage (KRI) et krigeage + Stratification Contrôlée (K-SC). Les résultats pour θ_2 se trouvent dans la Figure 45.

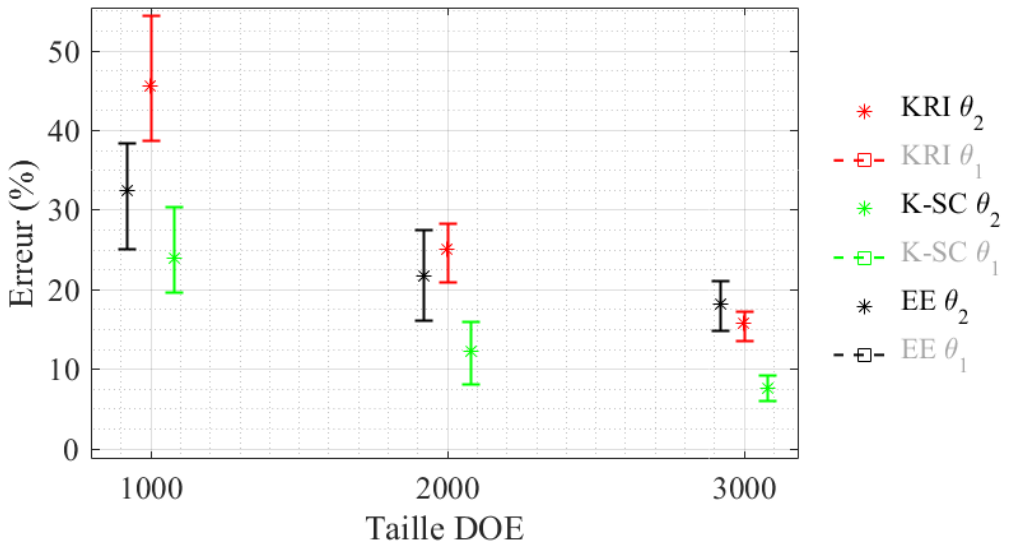

Figure 45 : Modèle S11, estimation de $q_{1\%}$. Erreur relative absolue maximale (θ_2) par estimation empirique (EE), krigeage (KRI) et krigeage + Stratification Contrôlée (K-SC). Les résultats pour θ_1 se trouvent dans la Figure 44.

3. Modèle PCB : Méthodologie et résultats

Le traitement de ce cas test est différent de l'exemple précédent (S11). Plutôt que de nous focaliser à nouveau sur l'estimation des statistiques d'erreur qui produiront les mêmes conclusions que pour le cas précédent, nous nous rapprochons d'une situation plus réaliste. En effet, en pratique, nous ne disposerons que d'un seul DOE de taille *n* éventuellement variable si on ajoute progressivement de nouveaux calculs du modèle vrai. L'estimation de l'erreur d'estimation peut alors être observée en fonction de cette taille de DOE.

L'algorithme mis en œuvre pour ce cas d'application suit donc uniquement les trois premières

étapes présentées en début de section V. Cependant, contrairement au modèle S_{11}, nous effectuons ces étapes qu'une seule fois (et non plus 32) pour chaque taille *n* choisie. Cette étude de cas est plus proche d'un cas d'application à la différence majeure près que nous nous autorisons à calculer une erreur d'estimation pour une taille de DOE donnée.

L'erreur pour une taille du DOE donnée est en réalité une variable aléatoire ayant potentiellement de grandes fluctuations statistiques comme nous l'avons vu avec le modèle S_{11}. De plus, il n'est pas possible, à proprement parler, de regrouper certaines réalisations provenant de DOE de tailles différentes pour en déduire une statistique parce qu'elles ne sont pas échantillonnées à partir de la même distribution. Cependant nous supposons que pour les tailles proches de DOE la distribution reste relativement identique. Nous avons utilisé une fenêtre coulissante moyenne de taille 30 pour lisser le bruit des erreurs avec la fonction Matlab *movmean*.

Les erreurs d'estimation de $q_{90\%}$, $q_{99\%}$ *et* $q_{99.9\%}$ sont tracées en Figure 46, Figure 47 et Figure 48 respectivement. Les points représentent la réalisation des erreurs tandis que les lignes pleines représentent les erreurs lissées au moyen de la fenêtre glissante évoquée précédemment.

Pour chaque tracé et chaque méthode, les erreurs lissées convergent vers 0, et les réalisations sont moins dispersées à mesure que la taille du DOE augmente.

Les conclusions que nous pouvons tirer de l'estimation de $q_{90\%}$ à partir de la Figure 46, sont les suivantes en fonction de la valeur de *n* :

- $n \leq 100$: aucune des trois méthodes ne fonctionne bien.
- $n > 100$: l'erreur moyenne de K-SC est plus proche de zéro que l'erreur moyenne de KRI, mais présente de rares valeur très élevées. Le KRI présente un faible biais négatif. EE, bien que sans biais, présente une variance élevée.

Les conclusions que nous pouvons tirer de l'estimation de $q_{99\%}$ à partir de la Figure 47 sont les suivantes en fonction de la valeur de *n* :

- $n \leq 200$: aucune des trois méthodes ne fonctionne bien.
- $n > 200$: l'EE procure l'estimation la moins fiable. Le KRI donne une très bonne estimation quoique très faiblement biaisée. L'estimation de K-SC n'est pas biaisée, avec une faible variance. K-SC semble donc surpasser légèrement le KRI.

Enfin, l'observation des résultats d'estimation du quantile $q_{99.9\%}$ tracés à la Figure 48 montre que la limite des domaines précédemment cités (100 et 200) est décalée vers $n \approx 350$. Il apparaît que pour un quantile aussi extrême, le K-SC surpasse nettement l'EE et le KRI.

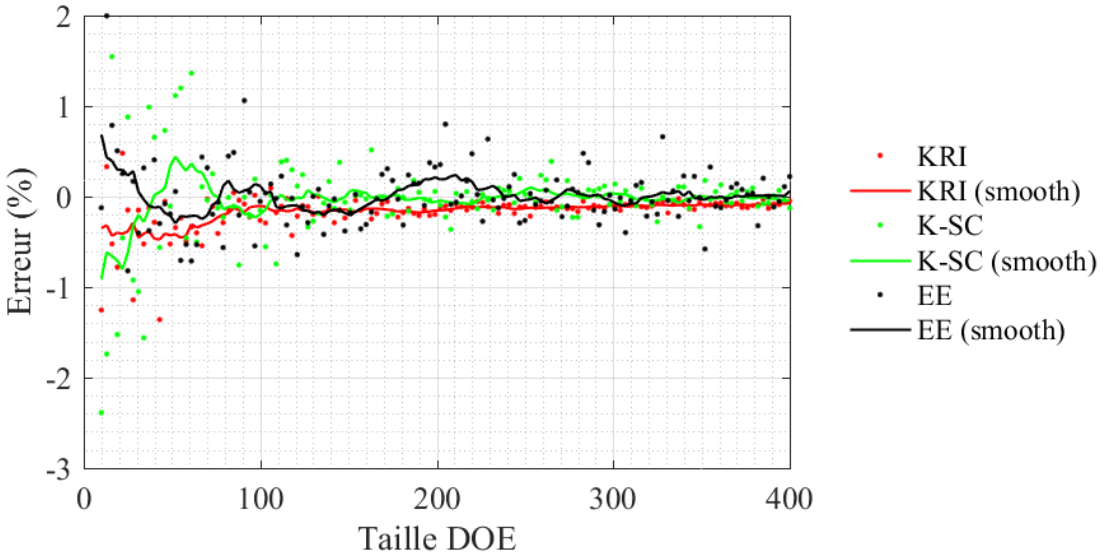

Figure 46 : Estimation du rayonnement des PCB $q_{90\%}$: erreur relative (%) par estimation empirique (EE), krigeage (KRI) et krigeage + Stratification Contrôlée (K-SC). Les courbes « smooth » sont obtenue par filtrage des points (moyenne par fenêtre glissante).

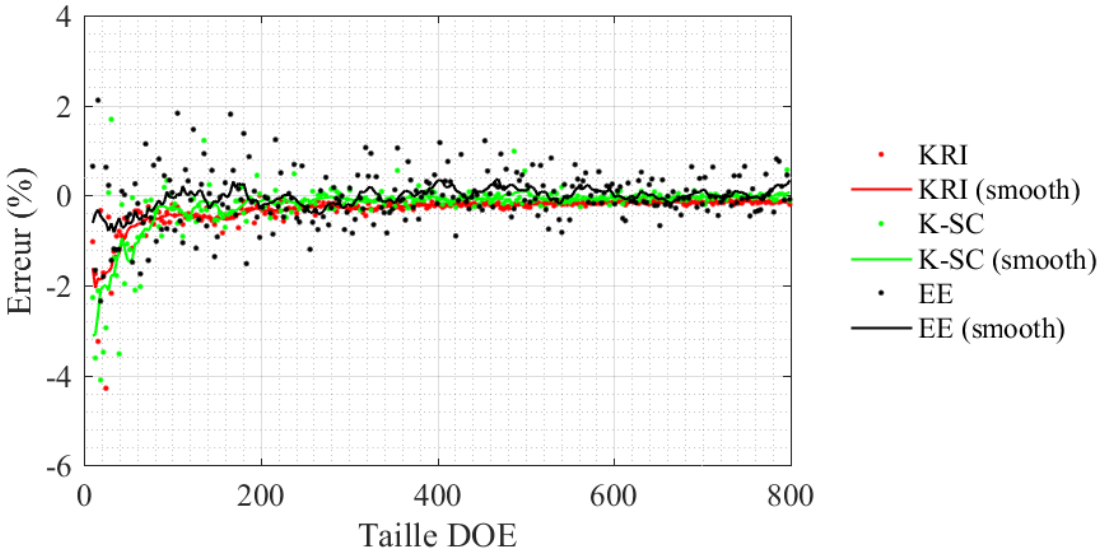

Figure 47 : Estimation du rayonnement des PCB $q_{99\%}$: erreur relative (%) par estimation empirique (EE), krigeage (KRI) et krigeage + Stratification Contrôlée (K-SC). Les courbes « smooth » sont obtenue par filtrage des points (moyenne par fenêtre glissante).

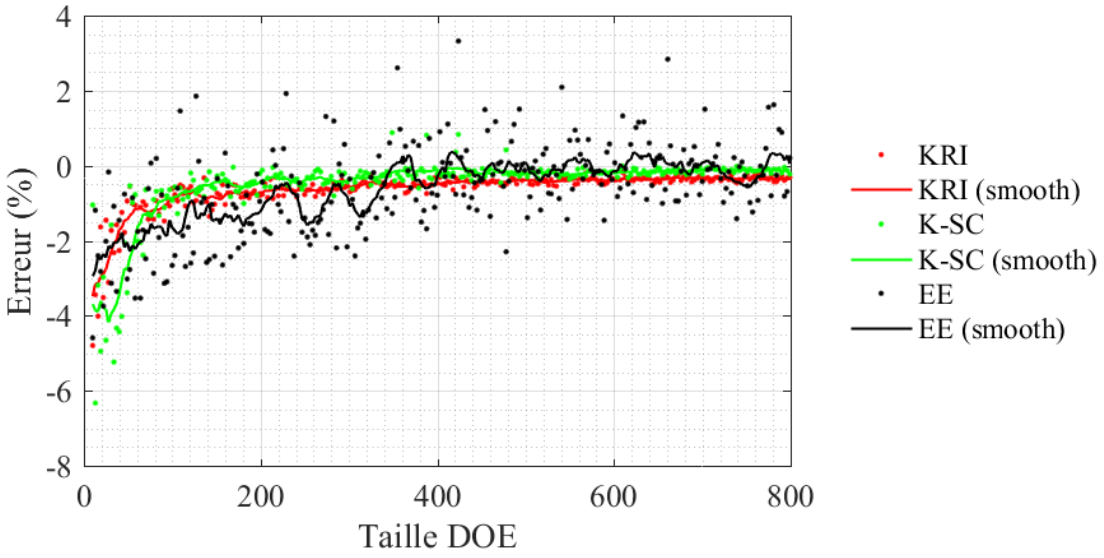

Figure 48 : Estimation du rayonnement des PCB $q_{99,9\%}$: erreur relative (%) par estimation empirique (EE), krigeage (KRI) et krigeage + Stratification Contrôlée (K-SC). Les courbes « smooth » sont obtenue par filtrage des points (moyenne par fenêtre glissante).

4. Conclusion sur l'intérêt de la stratification contrôlée

Nous avons introduit la stratification contrôlée basée sur le krigeage. Nous avons procédé à deux études. La première visait à choisir un SM pouvant être utilisé par la stratification contrôlée. La deuxième visait à étudier la valeur ajoutée de la SC par rapport à l'utilisation du SM seul.

Une première étude (section IV) a d'abord comparé les performances de plusieurs SMs dans la zone des valeurs extrêmes. Pour les deux modèles étudiés (PCB et S_{11}), le KRI ainsi que le PCE_KRI ont obtenu les meilleurs résultats. Même si le NN est aussi performant, il ne fournit pas d'erreur locale contrairement au KRI et au PCE_KRI. De plus, le temps de calcul de la partie PCE du PCE_KRI retarde considérablement sa construction. Cette première étude a donc permis de sélectionner le krigeage comme le meilleur compromis en vue d'une hybridation avec méthode de SC.

La deuxième étude (section V) visait à tester l'intérêt de rajouter la technique de stratification contrôlée au krigeage. Nous avons montré que la méthode K-SC surpasse l'approche de KRI seul, pour une estimation de quantiles extrêmes. En effet, le K-SC devient plus pertinent que le KRI lorsqu'on estime $q_{1\%}$ du modèle S_{11} (forte non-linéarité) et $q_{99.9\%}$, du modèle PCB (non-linéarité peu marquée). La K-SC surpasse le KRI d'autant plus que la non-linéarité du modèle (dans la région des valeurs extrêmes) augmente et que l'on recherche des quantiles de plus en plus extrêmes. Enfin, pour les modèles de très grandes dimensions ou pour de grand DOE, le temps de construction du KRI peut devenir important. Or la K-SC nécessite de construire un KRI avec un DOE deux fois plus petit que KRI seul. Le temps de construction de K-SC est donc plus rapide que celui consacré au KRI seul.

Dans le contexte de l'analyse des risques de l'IEMI et de la détermination du risque de défaillance à faible probabilité, le K-SC est donc pertinent pour parvenir à une approximation acceptable avec un DOE de taille limitée et apporte une réelle plus-value en terme de performances par rapport au KRI seul.

Jusqu'à présent, les modèles utilisés (PCB et S11) étaient peu coûteux car analytiques. C'est pourquoi nous n'avons pas cherché en priorité à minimiser le nombre d'appel à ce modèle, i.e. la ou les tailles de DOE à tester. Or en pratique, il ne sera pas possible de tester toutes les tailles possibles, ni même d'évaluer a priori la taille du DOE requise pour estimer correctement le quantile recherché ou plus généralement l'extrémité de la distribution de la contrainte. C'est pourquoi dans la section

suivante nous proposons une solution pratique implémentant la technique de K-SC dans le contexte où la question de la taille du DOE et donc du budget de simulation, est centrale.

VI. Gestion du budget de simulation lors de l'estimation d'un quantile extrême par la méthode K-SC

Les modèles de simulation CEM couramment utilisés sont de type *full-wave*. Bien que très précis, ils sont cependant très long à simuler. En effet, le calcul d'une seule réalisation de sortie peut durer plusieurs minutes pour des géométries simples (e.g. une piste PCB) jusqu'à plusieurs heures, voire plusieurs jours pour des systèmes complets (e.g. un véhicule). La bonne gestion du budget de simulation et donc de la taille du DOE est cruciale en pratique. Cette section propose une implémentation pratique de la K-SC où la question de la gestion du budget de simulation est centrale. L'objectif de cette approche est d'évaluer si le budget de simulation disponible, que l'on cherche le plus souvent à minimiser, est suffisant pour obtenir une bonne estimation du quantile visé.

Évidemment, la performance de K-SC (ou de toute estimation statistique) s'améliore à mesure que la taille du DOE augmente. Cependant, pour une taille de DOE donnée, on peut difficilement prédire la variance de l'estimateur. L'algorithme que nous présentons ci-après est donc nécessaire pour déterminer la taille de DOE suffisante pour estimer correctement le quantile visé.

1. Présentation de l'algorithme de gestion du budget de simulation

Le but de cet algorithme est d'estimer le taux d'amélioration de la performance SM en fonction de la taille du DOE en examinant le changement relatif de la moyenne ou de l'écart-type de l'estimateur du quantile. Par exemple, si l'amélioration est insignifiante, selon l'utilisateur, l'augmentation du budget de simulation n'est pas pertinente. Au contraire, si le changement relatif est important, toujours selon l'utilisateur, il est utile d'augmenter le budget de simulation.

Cet algorithme est principalement basé sur l'observation des performances du KRI. L'approche K-SC estime un quantile grâce à la SC sur la base d'un KRI. Par conséquent, l'estimation du quantile effectuée avec la SC s'améliore nécessairement en même temps que celle du KRI.

a. Description de l'algorithme

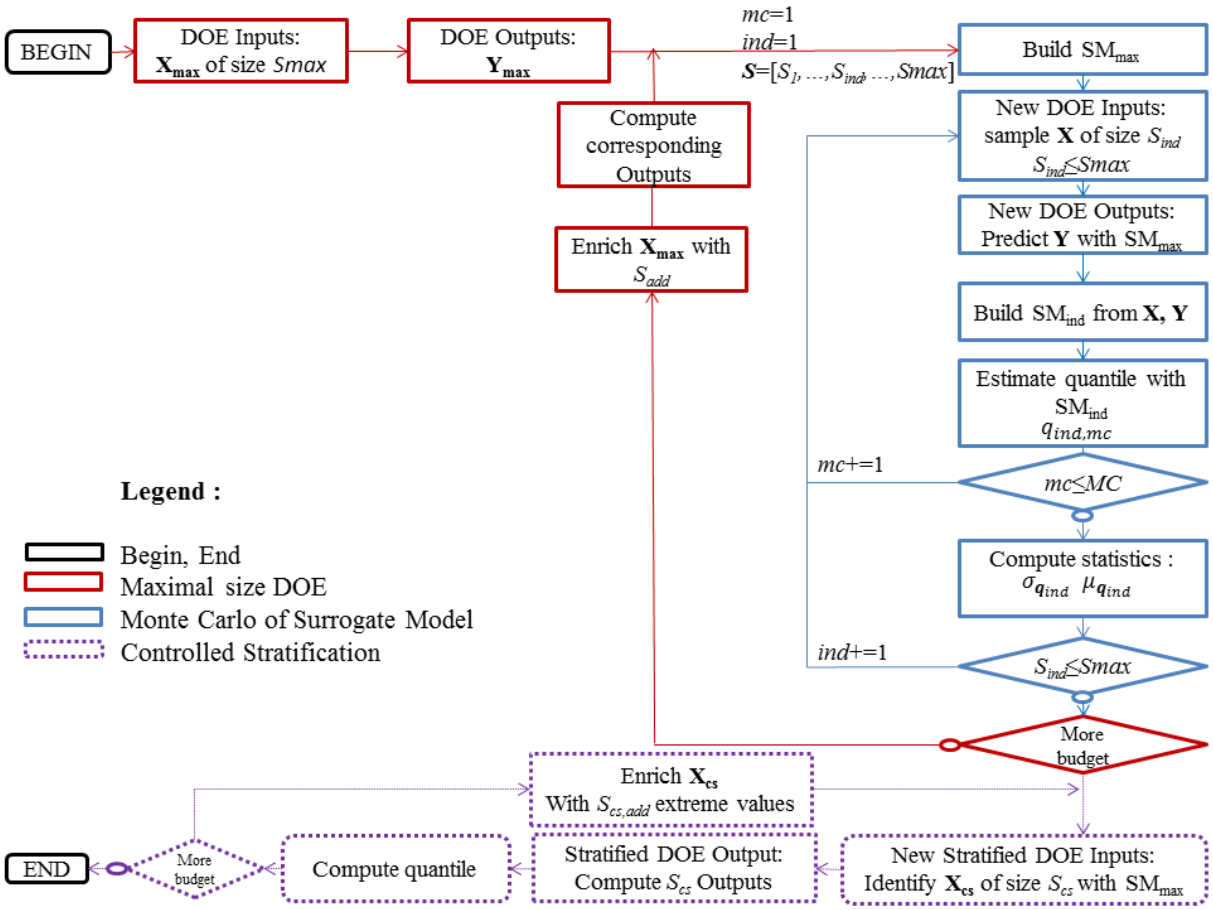

Figure 49 : Schéma de l'algorithme d'implémentation pratique de K-SC.

L'algorithme est représenté à la Figure 49. Comme l'indique le code couleur, il est composé de trois parties principales :

a. Gestion du DOE de taille maximale initiale *Smax* (en rouge)
b. Construction de différents SMs (krigeage) et simulations Monte Carlo associées (en bleu)
c. Stratification contrôlée (en violet).

Il est important de noter que le modèle coûteux n'est appelé que dans les parties 1 et 3. Par conséquent, le nombre total de fois (N) que le modèle réel est utilisé est : $Smax+S_{add}$ dans la partie 1 plus $S_{cs}+S_{cs,add}$ dans la partie 3. $Smax$ et S_{cs} correspondent au budget initial de simulation pour la construction du SM d'une part et la mise en œuvre de la SC d'autre part. S_{add} et et $S_{CS,add}$ constituent d'éventuelles simulations additionnelles.

L'initialisation de l'algorithme repose sur un DOE dont la taille maximale, *Smax*, doit être soigneusement choisie. A partir de ce DOE initial, un SM est alors construit. Dans le cas de la méthode K-SC, il s'agit du krigeage, choix que nous avons effectué au regard de notre analyse en section IV. Cependant, l'algorithme que nous présentons est généralisable à toute forme de SM. Ce SM est ensuite utilisé pour prédire la sortie de nouvelles entrées indépendantes, formant ainsi un nouveau DOE, cette fois de plus petite taille (S_{ind}). Pour chaque DOE plus petit, un nouveau SM est construit et utilisé pour estimer le quantile visé. Tandis que le modèle réel n'est appelé que pour créer le premier DOE de taille maximale, l'estimation de quantile est dès lors disponible pour les DOEs de plus petites tailles car il repose uniquement sur la construction successive de plusieurs SMs à faibles coûts de calcul. Par conséquent, les fluctuations de l'estimation du quantile α peuvent être évaluées à

un coût très réduit. Cette information est utile pour indiquer si le budget de simulation est suffisamment important. Une fois que l'estimation du quantile de la SM est jugée suffisamment stable, on utilise le SM le plus précis, i.e. celui construit à la taille maximale pour obtenir l'estimation du quantile final. A notre connaissance, outre la méthode K-SC, cette implémentation de celle-ci n'a encore jamais été introduite dans la littérature.

Nous allons à présent détailler chacune des trois parties de l'algorithme.

b. Gestion du DOE de taille maximale initiale Smax (section rouge de la Figure 49)

Un DOE de taille maximale autorisée *Smax*, est généré : l'échantillon d'entrée $\mathbf{X_{max}}$ est créé à partir d'un échantillonneur LHS [138]. Ensuite, le modèle coûteux est appelé *Smax* fois afin de calculer l'échantillon de sortie $\mathbf{Y_{max}}$ correspondant.

La performance de la stratification contrôlée dépend de la capacité de prédiction des réalisations extrêmes du SM. En effet, le SM est chargé d'identifier les bons candidats de l'espace d'entrée qui devraient être dans la strate ciblée. Cependant, un tel SM doit être construit avec un DOE qui contient au moins une réalisation extrême pour que sa prédiction commence à être fiable dans la queue de la distribution de sortie. En effet, si le SM n'a jamais échantillonné ne serait-ce qu'une seule valeur extrême lors de sa construction, sa prédiction constitue alors une extrapolation qui ne peut pas être considérée comme fiable.

Afin de garantir l'obtention probable d'au moins une valeur extrême dans le DOE de taille *N*, nous proposons une approche simple basée sur la distribution binomiale négative [139]. En effet, nous nous situons alors dans le contexte d'une expérience de Bernoulli à *N* essais. A chaque essai, un succès est comptabilisé si la réalisation de sortie est supérieure/inférieure (queue droite/gauche de la distribution) au quantile α visé. La probabilité de succès est donc de $p = min\,(\alpha, 1 - \alpha)$. Nous souhaitons donc quantifier la probabilité de compter au moins une réalisation de sortie extrême dans un DOE de taille *N*. Soit *K* une variable aléatoire comptant le nombre d'échecs *k* (i.e. les réalisations inférieure/supérieure au quantile de probabilité α) avant d'observer *n* essais réussis (i.e. supérieurs/inférieurs au quantile de probabilité α). La variable *K* suit alors une distribution binomiale négative de paramètre *p* et *N*, dont la *cdf* est :

$$F_K(k) = 1 - p^{k+1} \sum_{i=0}^{N-1} \binom{k+i}{i} (1-p)^i. \tag{58}$$

La *cdf* inverse, également connue sous le nom de fonction quantile, d'une telle distribution est utilisée pour calculer la taille minimale de DOE désignée par N_{min} (à un intervalle de confiance près fixé à 95 %) pour observer au moins *n* valeurs extrêmes :

$$N_{min}(n,p) = F_K^{-1}(0.95|n,p) + n. \tag{59}$$

Nous avons utilisé la fonction Matlab *icdf* pour calculer l'inverse de la *cdf* $F_K(k)$. Certaines valeurs numériques de N_{min} sont données dans le Tableau 16 en fonction de la probabilité des quantiles α visés et du nombre de valeurs extrêmes *n* souhaité. Par exemple, afin de garantir avec une confiance de 95 %, qu'au moins une réalisation sera inférieure au quantile 1 %, il faut un DOE d'une taille minimale égale à 299.

Tableau 16 : Estimation du nombre de réalisations nécessaires afin d'obtenir au moins n événements extrêmes en fonction de la probabilité p (%) d'événements pour un intervalle de confiance de 95 %.

p \ *n*	1	2	3	4	5	6	7	8
1	299	473	628	773	913	1049	1182	1312
2	149	236	313	386	456	523	590	655
3	99	157	208	257	303	348	392	436
4	74	117	156	192	227	261	294	326
5	59	93	124	153	181	208	234	260
6	49	78	103	127	150	173	195	217
7	42	66	88	109	129	148	167	185
8	36	58	77	95	112	129	146	162
9	32	51	68	84	100	115	129	143
10	29	46	61	76	89	103	116	129

Une stratégie d'enrichissement permet ensuite d'ajouter plus de réalisations au DOE sans altérer les propriétés LHS du DOE [140] et, surtout, sans gaspiller les échantillons précédemment calculés. Cet enrichissement est prévu lorsque les performances du SM sont estimées insuffisantes à la taille du DOE considérée. Cette évaluation est réalisée selon la procédure donnée ci-dessous.

c. Construction de différents SMs (krigeage) et simulations Monte Carlo associées (section bleue de la Figure 49)

A partir de l'échantillon maximal de taille *Smax* d'entrées et de sorties du DOE, un SM est construit à la taille maximale. Ce SM est noté SM_{max}. Deux indices sont également initialisés (*ind* et *mc*) comme le montre la Figure 49. L'indice *ind* identifie la taille du DOE en cours de test. Chaque taille testée est inférieure ou égale à la taille maximale. Elles sont regroupées dans un vecteur de tailles $\boldsymbol{S}=[S_1,...,S_{ind},...,Smax]$. Le choix des tailles à tester n'est pas critique. La stratégie la plus simple consiste à fixer un ensemble de tailles possibles, jusqu'à la taille maximale abordable choisie dans le Tableau 16. L'indice *mc* sert de compteur dans la boucle de simulation de Monte Carlo.

Le SM sélectionné est un krigeage implémenté dans la *toolbox* « UQLab » comme décrit dans la section II.6.

Un nouvel échantillon du vecteur **X** de taille S_{ind} est fourni à partir de la même distribution sous-jacente associée à chaque variable d'entrée, en utilisant le même échantillon LHS. Le SM (et non le modèle initial) est appelé (à faible coût de calcul) pour prédire les sorties **Y** correspondantes. En conséquence, un nouveau DOE est créé sans qu'il soit nécessaire d'exécuter à nouveau le modèle complexe. A partir de ce nouveau DOE de taille S_{ind}, un autre SM est construit, noté SM_{ind}, pour estimer le quantile ciblé.

Ce processus de génération d'un DOE de taille S_{ind} et de construction d'un nouveau SM (SM_{ind}) est répété *MC* fois afin d'effectuer une analyse statistique à partir des MC quantiles associés à la taille de l'échantillon S_{ind}. A partir de cet échantillon de quantiles, nous calculons sa moyenne μ_{qind} et son écart-type σ_{qind}. Le nombre *MC* correspond au nombre total de boucles de Monte Carlo et dépend des statistiques d'intérêt. Comme nous n'avions besoin que d'une estimation approximative des deux premiers moments, nous avons choisi *MC*=20.

Finalement, nous obtenons un échantillon multivarié de quantiles estimés (un échantillon de

quantiles pour chaque taille testée).

Deux informations peuvent être déduites de l'échantillon multivarié :

1. Détecter si la taille maximale du DOE choisie a permis d'éviter la zone des petits échantillons où les performances sont catastrophiques. Si la taille maximum de DOE a été soigneusement choisie comme discuté précédemment, il est peu probable que cela se produise.

2. Evaluer le taux d'amélioration du SM à partir de l'évolution des statistiques d'estimation du quantile en fonction de la taille du DOE. Si cette variation tend à être insignifiante, la taille maximale choisie du DOE convient. Au contraire, si les variations de l'estimation sont importantes, cela constitue une forte indication que l'enrichissement du DOE, en utilisant le modèle complexe, sera rentable.

d. Stratification contrôlée (section violette de la Figure 49)

Un nouveau DOE stratifié est créé pour SC à partir du krigeage initial (appelé plus haut le *SMmax*) construit à la taille maximale, *Smax*. La stratification de l'espace de sortie avec des quantiles adéquats est réalisée, et les réalisations d'entrées correspondantes sont identifiées. Nous avons choisi quatre strates définies par (47) et une stratégie d'allocation uniforme définie par (48).

Le modèle initial est ensuite appelé avec les réalisations d'entrée identifiées pour calculer les sorties correspondantes. Enfin, la distribution empirique pondérée des sorties est déduite avec (49) et le quantile est calculé selon (26).

Afin d'améliorer la stabilité de l'estimation du quantile par la SC sans avoir à calculer un nouveau DOE, la même stratégie d'enrichissement est utilisée. A partir du même *SMmax*, nous ajoutons quelques réalisations extrêmes prédites (affectées aux deux dernières strates). La distribution pondérée empirique est mise à jour pour calculer une autre estimation du même quantile. Ce processus est répété plusieurs fois.

2. Validation de l'algorithme sur les modèles analytiques

a. Description de la méthodologie

L'algorithme décrit dans la Figure 49 est en grande partie basé sur une simulation de Monte Carlo des statistiques de l'estimation d'un quantile réalisée avec le SM en fonction de la taille de l'échantillon / DOE, i.e. la section bleue du schéma de la Figure 49. Il utilise un même DOE initial et unique de taille maximale. Dans cette section, nous allons vérifier que les performances ne sont pas trop dépendantes du tirage aléatoire associé à ce DOE initial.

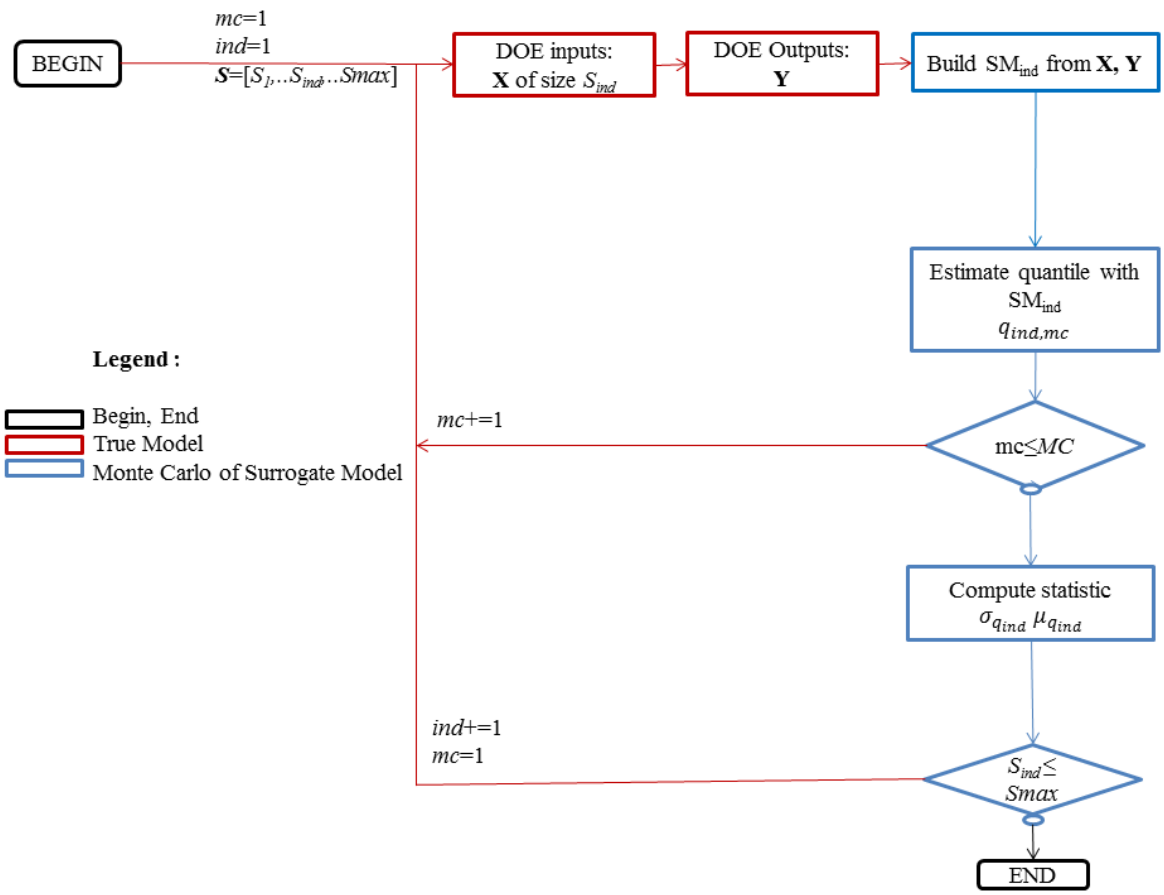

Figure 50 : Diagramme de l'algorithme pour valider partie SM de la Figure 49 (aussi en bleu).

Pour ce faire, l'algorithme de la Figure 49 est exécuté 10 fois (de BEGIN à END). Cependant, les sections concernant l'enrichissement du DOE et la SC de la Figure 49 ne sont pas appliquées pour cette analyse spécifique. De plus, nous fournissons une pseudo-référence à l'aide d'une simulation de Monte Carlo telle que décrite à la Figure 50 en choisissant *MC*=100. Afin d'établir cette pseudo-référence, pour chaque taille définie dans le vecteur $\boldsymbol{S} = [S1, S2, \ \ldots, Smax]$, un DOE est généré et le SM est construit à partir de l'échantillon complet. L'opération est renouvelée MC fois pour évaluer une statistique d'estimation associée à chaque taille de DOE : $S_1, \ldots, Smax$.

Pour réaliser cette étape de validation, nous utilisons les deux modèles analytiques déjà présentés. Le premier modèle est le coefficient de réflexion (en linéaire) d'un circuit RLC série à 4 entrées aléatoires. Il présente un comportement non linéaire en raison de sa nature résonnante. Le deuxième modèle est le champ rayonné émis par une trace de PCB avec 11 entrées aléatoires.

b. Modèles PCB et S11 : Résultats

i. Résultats de validation avec le modèle S11

Rappelons que les 4 entrées aléatoires du modèle S_{11} sont : la fréquence, la résistance (*R*), le condensateur (*C*) et l'inductance (*L*). La fréquence est répartie uniformément de 100 à 900 *MHz*. La résistance, le condensateur et l'inductance sont répartis uniformément entre 90 % et 110 % de leurs valeurs nominales. Les valeurs nominales (*R=50 Ω, C=1.5 pF* et *L=67.5 nH*) sont choisies pour atteindre une adaptation parfaite à 500 *MHz*.

Pour ce problème spécifique, nous visons à identifier le quantile à 1 % qui correspond à de très faibles amplitudes du coefficient de réflexion S11 à la résonance du circuit.

La Figure 51 et la Figure 52 montrent la moyenne μ_q et l'écart-type σ_q du quantile à 1 % pour chacune des 10 exécutions du schéma de la Figure 49 en fonction de la taille du DOE. *Smax* est toujours choisi égal à 770 pour chacune des 10 exécutions. Des points de couleur permettent de distinguer les 10 exécutions différentes de l'organigramme de la Figure 49. La pseudo-référence (issue de l'algorithme de la Figure 50) apparaît comme une courbe noire.

La Figure 51 et la Figure 52 présentent deux régions distinctes. Tout d'abord, nous identifions une petite région de taille du DOE inférieure à 300, où le quantile à 1 % est surestimé avec un écart-type important et instable d'une série à l'autre. Cela correspond aux attentes du Tableau 16 et par conséquent au risque d'exploiter un échantillon de départ sans valeurs extrêmes. La deuxième région correspond à une convergence plutôt monotone. Les estimateurs de la moyenne et de la variance demeurent légèrement biaisés, mais la tendance à l'amélioration de l'estimation est confirmée et s'approche de l'estimation fournie par la pseudo-référence. Par conséquent, l'algorithme proposé parvient très clairement à distinguer la région de petite taille. De plus, au-delà de cette région, le degré d'amélioration est bien estimé malgré un léger biais. A partir d'une observation certes subjective de ces résultats, il apparait que l'augmentation du DOE au-delà d'une taille de 300 est inutile dans le cas présent.

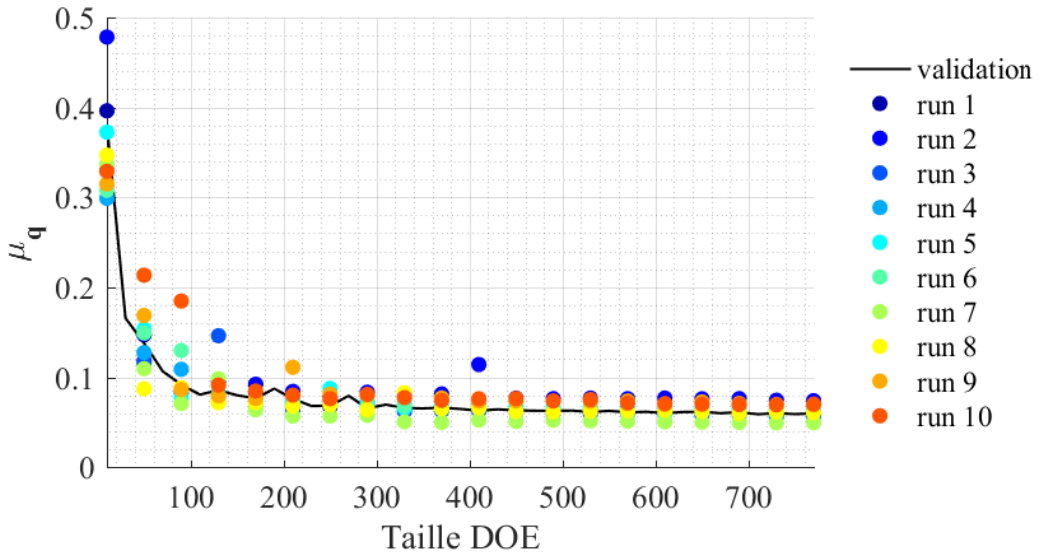

Figure 51 : Modèle S11 : évolution de la moyenne de l'estimation du quantile à 1 % en fonction de la taille du DOE. Les points représentent les évaluations de l'algorithme de la Figure 49. Chaque couleur correspond à une évaluation de l'algorithme. La courbe représente la pseudo-référence issue de l'algorithme en Figure 50 .

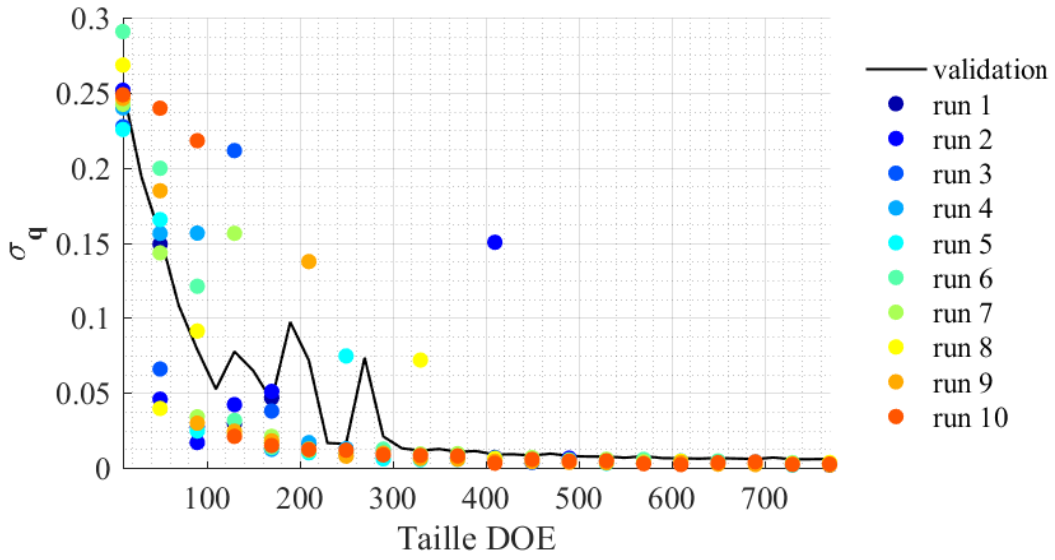

Figure 52 : Modèle S11 : évolution de l'écart type de l'estimation du quantile 1 % en fonction de la taille du DOE. Les points représentent les évaluations de l'algorithme de la Figure 49. Chaque couleur correspond à une évaluation de l'algorithme. La courbe représente la pseudo-référence issue de l'algorithme en Figure 50 .

ii. Résultats de validation avec le modèle PCB

Rappellons que les 11 entrées aléatoires du modèle PCB sont : la fréquence (f_r), les caractéristiques géométriques (l'épaisseur (h), la largeur (W) et la longueur (l) du substrat), les caractéristiques électriques (permittivité (ε_r), source de tension (Vs), source d'impédance (Zs), charge (Zl)) et la position où le champ est mesuré (coordonnées r, θ et φ). Chaque entrée suit une distribution gaussienne centrée sur sa valeur nominale avec un écart-type relatif de 10 %. Les valeurs nominales sont choisies de manière à ce que la courbe de mesure apparaisse comme une ligne de transmission en quart de longueur d'onde : f_r=404 *MHz*, h=0,775 *mm*, W=0,51 *cm*, l=10,16 *cm*, εr=4,6, Vs=1 *V*, Zs=50 θ, Zl=1 Ω, r=3 m, $\varphi = \theta = 2\pi$. La sortie du modèle est l'amplitude du champ électrique rayonné.

Pour ce problème spécifique, nous visons à identifier le quantile à 99 % qui correspond aux valeurs extrêmes du champ rayonné, i.e. un risque du point de vue CEM.

Les résultats de cette étude de cas sont présentés à la Figure 53 pour la moyenne et à la Figure 54 pour l'écart-type. Comme nous visons un quantile extreme (de la queue supérieure i.e. α proche de 1), il est attendu que la moyenne soit sous-estimée. L'écart type reste important pour les échantillons de petite taille, mais semble stable. La convergence monotone apparaît pour des tailles plutôt petites, mais le taux de variation est faible au-delà d'une taille limite d'environ 200. De toute évidence, l'algorithme proposé réussit, à un coût limité (un seul DOE à la taille maximale disponible), à déterminer la tendance à la convergence de l'estimation de ce quantile en fonction de cette taille.

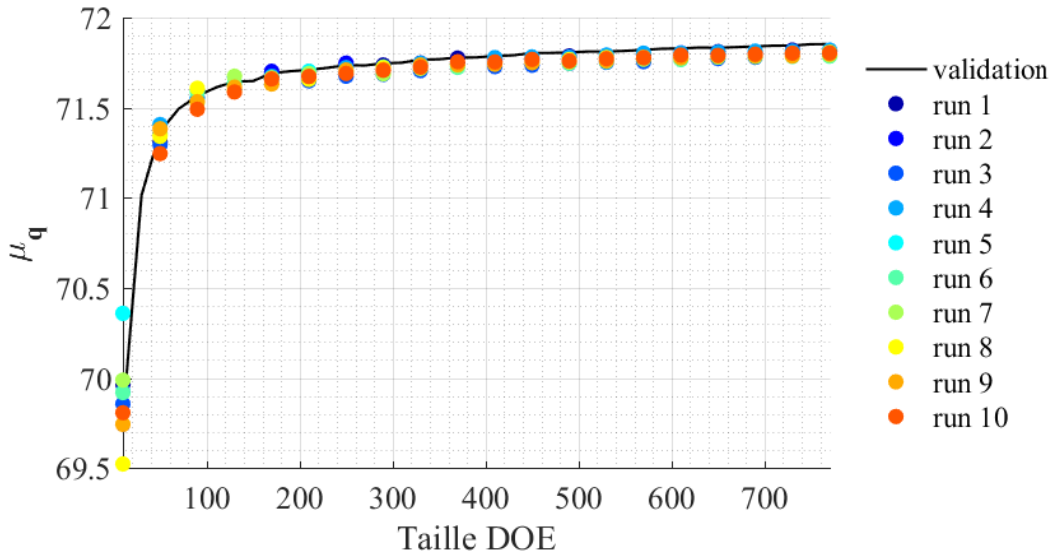

Figure 53 : Modèle PCB : évolution de la moyenne de l'estimation du quantile 99 % en fonction de la taille du DOE. Les points représentent les évaluations de l'algorithme de la Figure 49. Chaque couleur correspond à une évaluation de l'algorithme. La courbe représente la pseudo-référence issue de l'algorithme en Figure 50.

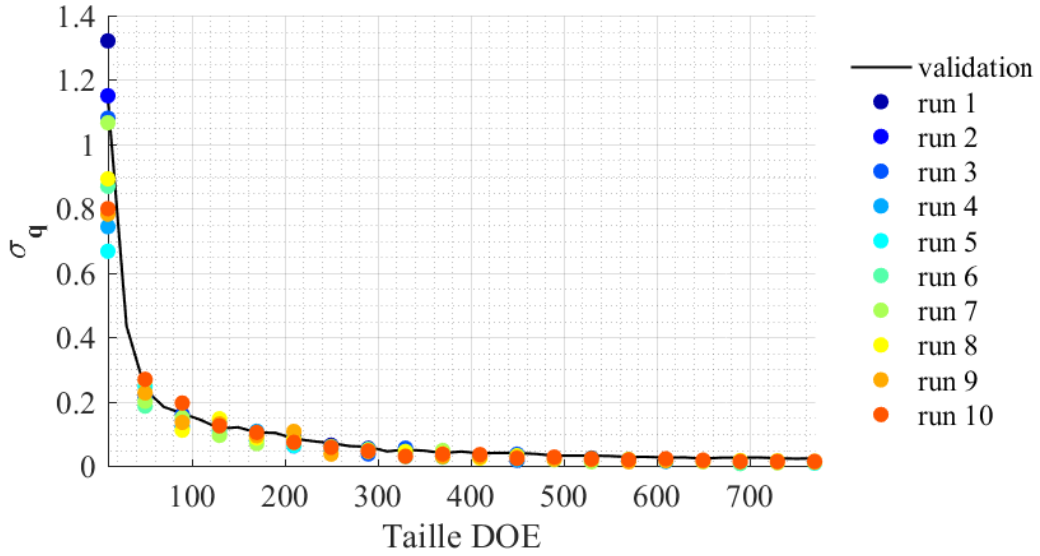

Figure 54 : Modèle PCB : évolution de l'écart type de l'estimation du quantile 99 % en fonction de la taille du DOE. Les points représentent les évaluations de l'algorithme de la Figure 49. Chaque couleur correspond à une évaluation de l'algorithme. La courbe représente la pseudo-référence issue de l'algorithme en Figure 50.

3. Cas d'application

Dans cette section, l'algorithme complet (y compris les enrichissements en réalisations, selon LHS, pour l'estimation du SM et en réalisations complémentaires pour la SC) est mis en œuvre dans un scénario d'essai d'émissivité rayonnée avec des incertitudes épistémiques sur les paramètres du modèle.

a. Présentation du modèle d'étude

Le modèle est une cavité rectangulaire comportant 2 ouvertures rectangulaires sur la face supérieure. Un courant circule sur la surface externe d'un fil conducteur plein à l'intérieur de la cavité. Les parois de la cavité ainsi que le fil sont faits de conducteurs électriques parfaits. Une vue 3D de la géométrie avec les entrées réglées à leur valeur nominale est représentée dans la Figure 55.

Le modèle est décrit par des paramètres. Certains sont considérés connus avec une valeur constante, d'autres sont incertains. Tous les paramètres d'entrée sont énumérés dans le Tableau 17. Les 16 variables aléatoires sont indépendantes. Elles suivent une distribution uniforme centrée sur leur valeur nominale avec une tolérance, arbitrairement fixée à +/- 10 % autour de cette dernière.

La sortie d'intérêt est l'amplitude maximale du champ électrique rayonné (E) sur une sphère à une distance d de la position de la cavité (Ω_c). La sortie nominale (i.e. la sortie lorsque les entrées sont réglées à leur valeur nominale) est égale à 0,01 V/m. Le quantile cible de E est le quantile à 99 %.

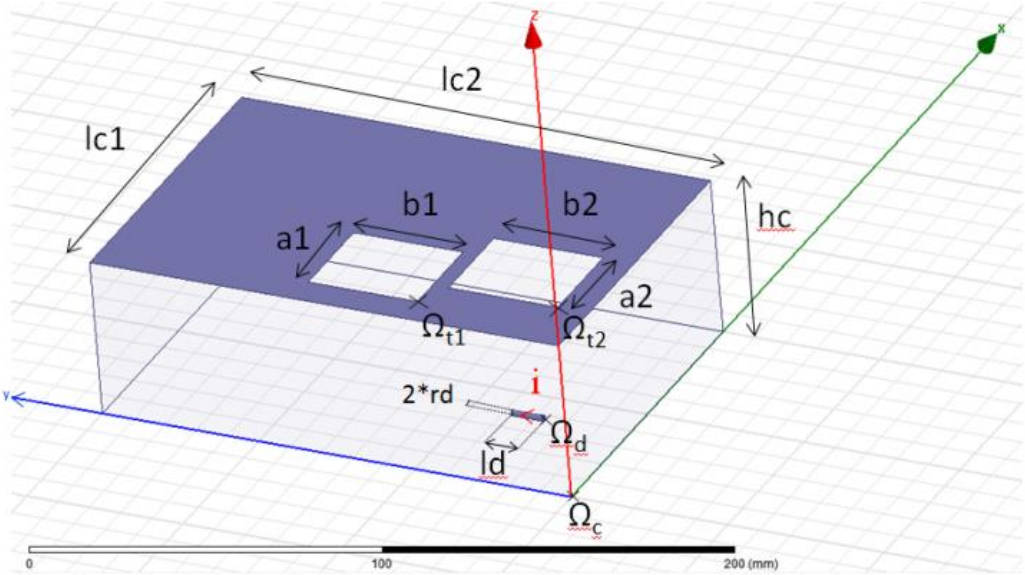

Figure 55 : Géométrie de la cavité comportant deux ouvertures et une source de courant sur un fil conducteur à l'intérieur de la cavité.

Tableau 17 : Entrées du modèle du fil à l'intérieur de la cavité percée.

Name	Description	Nominal value	Unit	Tolérance
xc	Position de la cavité (Ω_c)	0	*mm*	0
yc		0		
zc		0		
fr	Fréquence	500	*MHz*	
d	Distance par rapport à Ω_c	3	*m*	
i	Source de courrant	1	*A*	
ld	Longueur du fil	10	*mm*	10 %
rd	Rayon du fil	1		
xd	Position du fil (Ω_d)	15		
yd		15		
zd		15		
lc1	Longueur de la cavité	100		
lc2	Largeur de la cavité	150		
hc	Hauteur de la cavité	50		
xt1	Ω_{t1}: position de la 1ère ouverture	10		
yt1		50		
hc		*50*		
a1	Longueur de la 1ère ouverture	30		
b1	Largeur de la 1ère ouverture	35		
xt2	Ω_{t2} : Position de la 2ème ouverture	20		
yt2		10		
hc		*50*		
a2	Longueur de la 2ème ouverture	30		
b2	Largeur de la 2ème ouverture	35		

b. Echantillon de référence

Afin d'obtenir un échantillon de référence de la sortie, le logiciel HFSS est utilisé pour générer un échantillon d'entrées et calculer la sortie correspondante. La simulation a été arrêtée après 72 heures, ce qui a donné un DOE de taille 3225. Comme cela représente un nombre limité d'évaluations du modèle, le quantile à 99 % calculé à partir de l'échantillon de référence reste une estimation. Afin d'évaluer l'incertitude d'estimation de cette référence du quantile à 99 %, nous avons utilisé le Bootstrapping [54] qui permet de calculer les bornes de vraisemblance du quantile de référence. La Figure 56 montre un histogramme de la distribution de l'échantillon de référence. La moyenne de l'échantillon Bootstrap du quantile à 99 % est représentée sur la Figure 57 avec son intervalle de confiance à 95 % (quantile à 2.5 % et 97.5 %) de l'échantillon généré par Bootstrap).

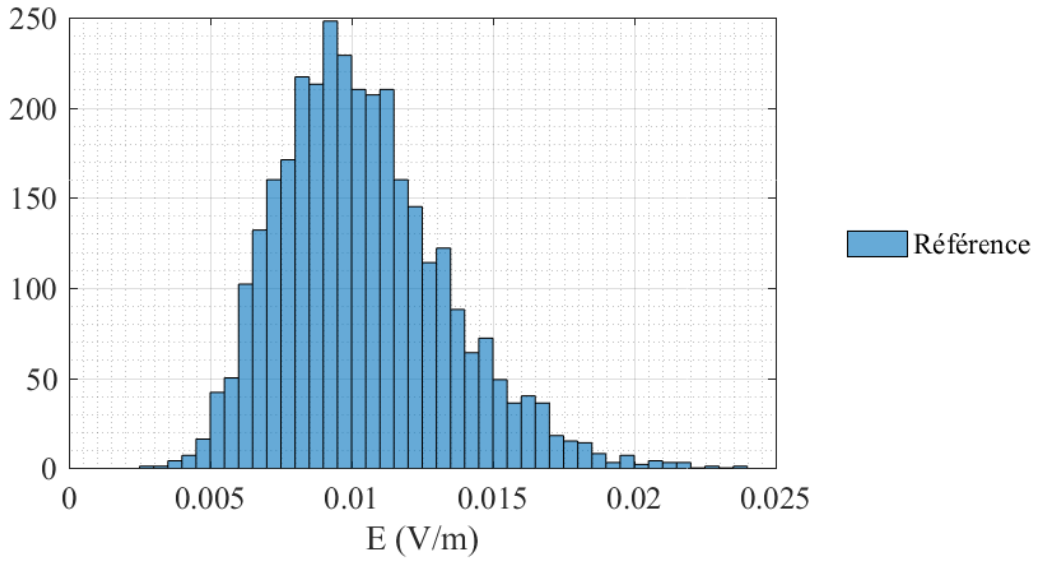

Figure 56 : Histogramme de la sortie de l'échantillon (champ électrique maximum sur une sphère à 3 mètres de distance de la cavité).

c. Résultats

i. Choix du DOE initial

La simulation d'une réalisation (entrées réglées à leur valeur nominale) dure 30 secondes. D'après le Tableau 16, la taille de DOE suggéré pour estimer le quantile 99 % est de 299 (pour n = 1). Sous l'hypothèse d'une dépendance linéaire entre la taille de DOE, le temps de simulation nécessaire serait de 2.5 heures. Pour cet exemple, nous supposons que cela reste acceptable.

ii. Simulation Monte Carlo du SM

Après le calcul du DOE (de taille *Smax*=299) avec le modèle HFSS (section de couleur rouge de la Figure 49), nous exécutons la section de simulation SM MC de notre algorithme (section de couleur bleue de la Figure 49) pour 10 tailles de DOE comprises entre 10 et 299. La moyenne du quantile est représentée sur la Figure 57 (en bleu pour les tailles de DOE comprises entre 10 et 299) avec des barres d'erreur (une largeur de barre est égale à 3.92 fois l'écart type).

Compte tenu de cette taille initiale de 299, la partie bleue de la courbe de la Figure 57 montre que le taux de variation de l'estimation tend à être limité au-dessus de 150 réalisations. Nous pouvons donc conclure qu'un DOE de 299 est suffisant pour obtenir une approximation raisonnable avant d'appliquer l'approche SC.

iii. Enrichissement du DOE

Étant donné que le modèle HFSS fonctionne assez rapidement, nous supposons tout de même que nous pouvons nous permettre l'enrichissement du DOE initial avant d'appliquer le CS. De retour au Tableau 16, nous choisissons d'augmenter la taille du DOE de sorte que nous trouvions probablement au moins 2 valeurs extrêmes de sortie, ce qui donne une nouvelle taille totale du DOE de 473 (n=2). Ainsi, nous ajoutons 174 réalisations de plus à l'ancien DOE, ce qui donne un nouveau DOE de taille 473.

La simulation Monte-Carlo du SM est ensuite réalisée avec différentes tailles allant de 299 à la nouvelle taille maximale de 473. La partie bleue de la courbe de la Figure 57 est maintenant complétée par une nouvelle section de couleur magenta avec ses barres d'erreurs correspondantes. La variation de

l'estimation en fonction de la taille est nettement plus lente. De plus, on observe une discontinuité pour une taille d'échantillon de 299 entre la courbe bleue et la courbe magenta. En effet, les résultats montrés en bleu proviennent tous d'un SM construit avec un DOE de 299 alors que les résultats montrés en couleur magenta sont prédits par le SM construit avec un DOE de 473. Une continuité parfaite aurait donc été surprenante.

Afin de confirmer si besoin était la stabilité de l'estimation, nous avons procédé à un deuxième et dernier enrichissement. Nous avons donc ajouté un autre ensemble de 155 réalisations, ce qui porte le budget total de simulation à 628 (correspondant à n=3 dans le Tableau 16). Le même type de simulation Monte-Carlo du SM est effectué avec des tailles allant de 473 à 628 et le résultat apparaît dans la section verte de la courbe de la Figure 57. Il y a également une légère discontinuité à 473. Il se confirme donc qu'un tel enrichissement est superflu dans ce cas.

iv. Estimation du quantile par Stratification Contrôlée

La SC est réalisée en association avec le SM construit à partir de 628 réalisations. La SC permet d'enregistrer 628 nouvelles réalisations pour accélérer la convergence des estimations du quantile. Par conséquent, le quantile est estimé avec la SC à partir d'un budget total de 1256 et est tracé dans un cercle rouge non rempli à la Figure 58.

Enfin, afin de vérifier la stabilité de l'estimation du quantile par la SC, quelques réalisations supplémentaires sont attribuées dans la dernière strate. Le quantile estimé par cette stratégie d'enrichissement SC est également tracé en cercles rouges pleins.

Les quantiles estimés dans la Figure 58 oscillent autour de la limite supérieure de confiance de référence. La variation est négligeable par rapport à la largeur de la borne de confiance de référence, ce qui confirme la pertinence de notre algorithme.

Le nombre total d'appels au modèle réel est de 1296, ce qui représente environ 40 % de la taille de l'échantillon de référence (3225). Ce nombre aurait pu être réduit davantage en évitant le dernier enrichissement. Le nombre de réalisations supérieures au quantile estimé était de 32 avec l'échantillon de référence, il est de 125 avec la méthode K-SC (nombre de réalisations supérieures à 0.01931071 i.e. le quantile estimé final). Ceci procure donc l'opportunité d'approcher de manière plus précise la distribution des valeurs extrêmes du phénomène observé et les relations existantes entre l'obtention de ces valeurs extrêmes et les valeurs associées des paramètres d'entrée.

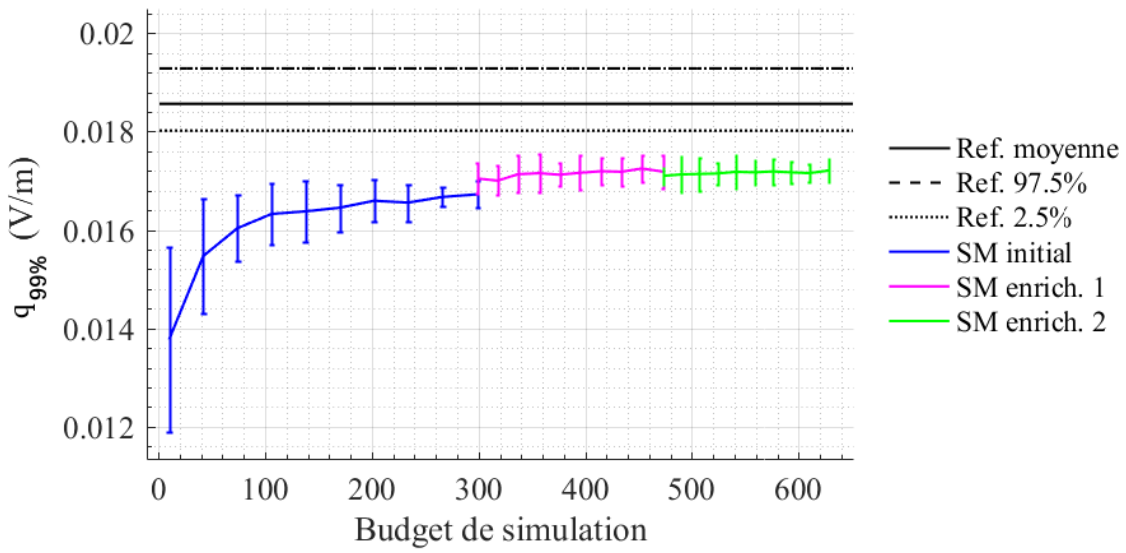

Figure 57 : Estimation du quantile 99% de E par l'échantillon de référence et le SM en fonction du nombre d'appel au modèle HFSS i.e. le budget de simulation.

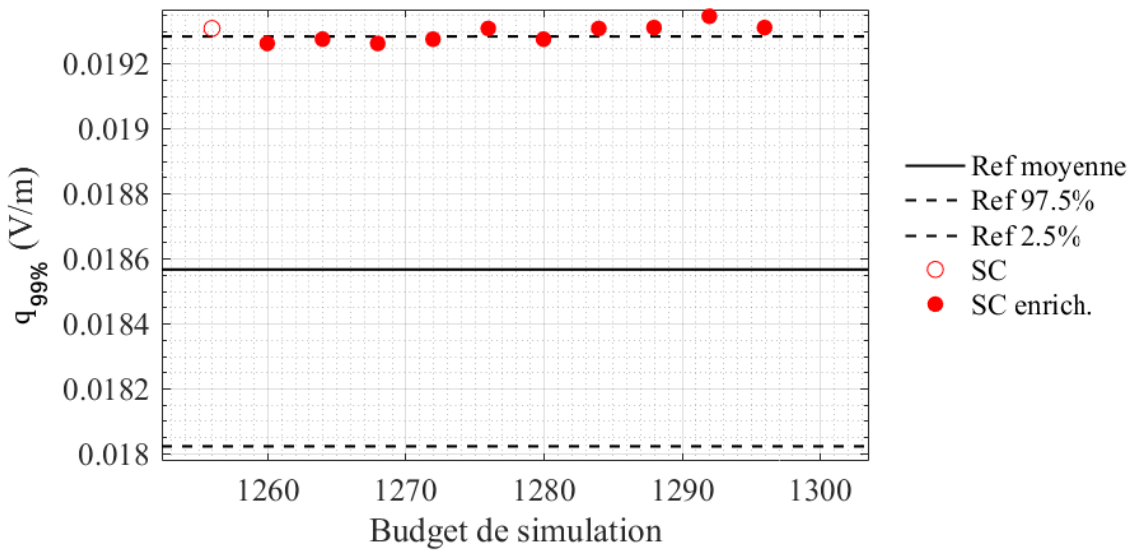

Figure 58 : Estimation du quantile 99% de E par la Stratification Contrôlée à la suite du SM en fonction du nombre d'appel au modèle HFSS i.e. le budget de simulation.

d. Analyse des valeurs extrêmes obtenues

Grâce au SM construit à partir des 628 réalisations et à l'application de la méthode K-SC, nous sommes en mesure d'identifier un large échantillon d'événements extrêmes. Nous utilisons cette information pour analyser les entrées qui sont susceptibles de les produire. Ces informations sont utiles pour améliorer la conception et éviter les situations à risques.

Dans la Figure 59, nous fournissons les histogrammes des réalisations ayant provoqué une valeur extrême. Les histogrammes sont normalisés à des fins de visualisation de sorte que la somme de la surface de chaque barre est égale à 1 (i.e. normalisation *pdf*). Les histogrammes bleus représentent les réalisations des entrées à l'origine de l'échantillon de sortie de référence de la Figure 56. Evidemment, chaque variable suit une distribution uniforme avec une variation de 10 % par rapport à sa valeur nominale comme prévu (voir Tableau 17). Les histogrammes de couleur orange représentent

la distribution des entrées qui ont généré des valeurs extrêmes (i.e. des valeurs supérieures au quantile estimé par la SC).

Une distribution uniforme n'est pas informative. C'est le cas par exemple des variables *a1* et *b1* (longueur et largeur de la première ouverture). Ces variables peuvent avoir un impact en conjonction avec d'autres variables mais cela ne peut pas être perçu à travers un tel histogramme. Une analyse de sensibilité globale avec des indices de Sobol d'ordre deux [140] serait nécessaire. Cependant, on peut conclure que ces variables ne déterminent pas à elles seules la probabilité d'obtenir des valeurs extrêmes d'émissions rayonnées.

D'un autre côté, si la répartition n'est manifestement pas uniforme, c'est une information très riche. C'est le cas pour *b2* et *yt1* (longueur de la deuxième ouverture et position de la première ouverture). Par exemple, une sortie extrême ne sera pas observée si *yt1*>48 *mm*, quelle que soit la valeur des autres entrées.

Par conséquent, la diminution de *b2* et l'augmentation de *yt1* sont les options à privilégier pour réduire le rayonnement hors de la cavité plutôt que de modifier *a1* et *b1*.

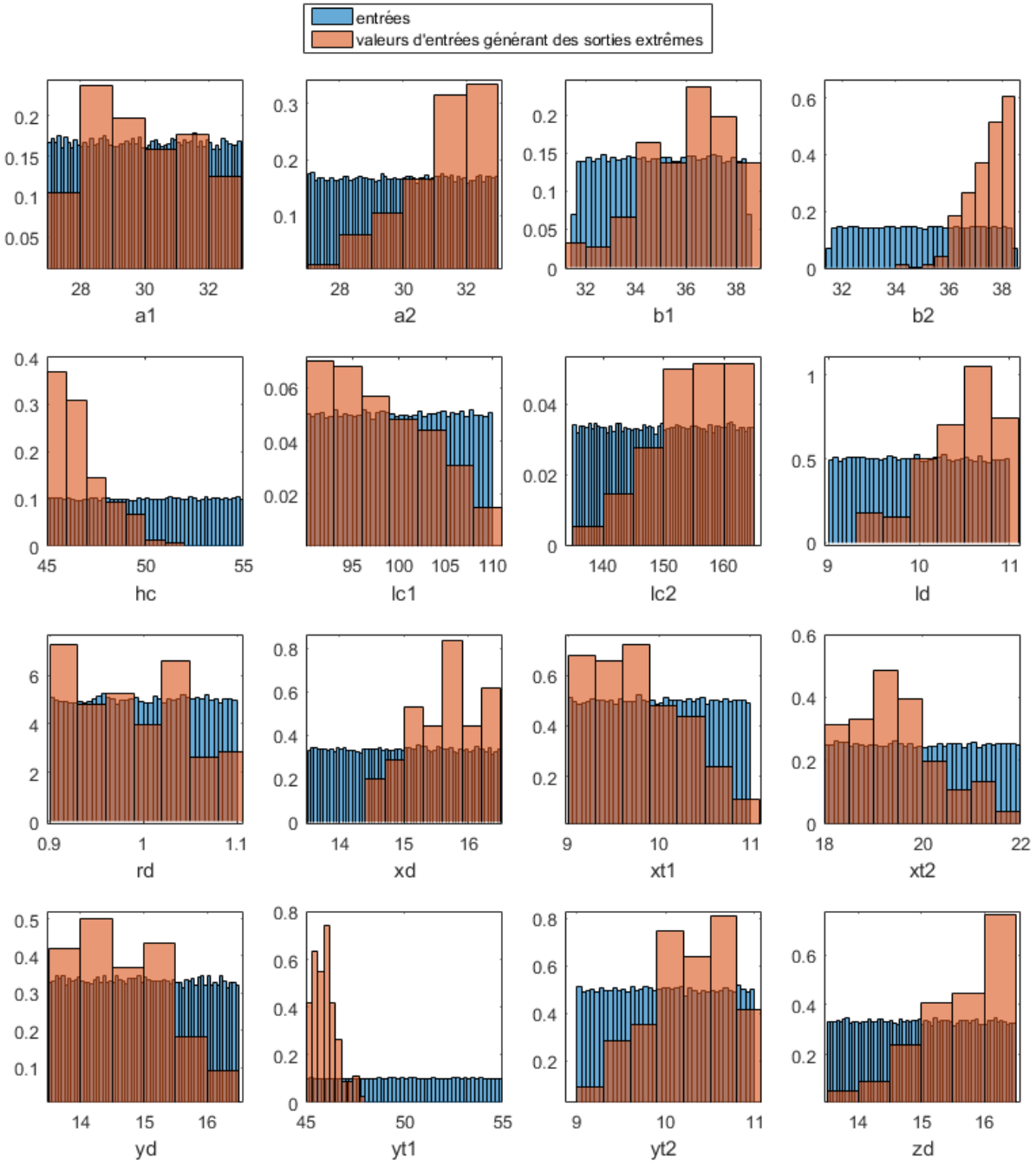

Figure 59 : Distribution des 16 variables d'entrées (bleu). Pour chaque variable, distribution des valeurs ayant provoquées des valeurs extrêmes de sortie (Orange).

4. Bilan sur l'intérêt pratique de cette implémentation de la Stratification Contrôlée

Les quantiles extrêmes fournissent des informations utiles pour l'analyse des risques CEM. Nous avons estimé les quantiles extrêmes de la distribution de sortie lorsque le modèle électromagnétique est coûteux à simuler. Nous avons introduit une mise en œuvre pratique d'une nouvelle approche appelée K-SC, qui associe le krigeage à la stratification contrôlée. L'algorithme proposé répond à la question pratique concernant le budget de simulation nécessaire pour atteindre une précision suffisante d'estimation. Cet algorithme fournit le taux d'amélioration au regard d'un budget de simulation réduit. La partie critique de l'algorithme a été validée sur des modèles analytiques. Ensuite, nous avons appliqué la procédure de calcul complète à une étude de cas CEM. Dans ce cas, nous avons montré qu'il surpassait nettement l'approche classique de Monte-Carlo. En particulier, K-SC fournit un nombre important de sorties extrêmes pour l'analyse de sensibilité ou l'analyse de

distribution des valeurs extrêmes. En plus de ses performances intéressantes, l'algorithme proposé peut être très facilement mis en œuvre. La partie krigeage est déjà implémentée dans la plupart des outils de programmation populaires (Matlab, Python et R). Le code Matlab de la SC est fourni en Annexes.

Cependant, pour les modèles de très grandes dimensions (quelques dizaines), la construction d'un modèle de substitution de krigeage n'est plus négligeable, de sorte que l'algorithme proposé devrait être modifié. Une perspective intéressante serait d'associer le krigeage à une méthode de régression de type *Partial Least Square* comme dans [136] pour des dimensions très élevées.

VII. Conclusion et pistes de réflexion

1. Conclusion

Ce chapitre visait à estimer la contrainte en cas d'IEMI en se focalisant sur l'estimation des valeurs extrêmes. La technique de stratification contrôlée est spécifiquement conçue dans ce but. Cependant, elle repose sur un modèle simple qui doit être associé au modèle vrai initial. La construction de ce modèle simple est une question restée ouverte, laissée à l'appréciation des utilisateurs.

Nous avons cherché précisément à adopter une démarche systématique pour la construction d'un tel SM qui pourrait assurer un contrôle efficace de la SC. Nous avons tout d'abord cherché le SM le plus à même de remplir cet objectif selon un critère de corrélation. Après comparaison de quelques SMs les plus courants, nous avons établi que le krigeage était le plus pertinent.

Nous avons ensuite cherché dans quelle mesure la SC apportait un gain de performance par rapport au krigeage utilisé seul, en vue de l'estimation de quantiles extrêmes. Nous avons montré au travers de deux modèles que plus la non-linéarité du modèle est importante, plus la SC est pertinente. Il en va de même si les valeurs de sortie visées sont de plus en plus rares.

Enfin, nous nous sommes intéressés à une implémentation pratique de notre méthode d'hybridation du krigeage avec la stratification contrôlée, baptisée K-SC. Nous avons proposé un algorithme permettant d'estimer la pertinence du choix a priori de la taille de l'échantillon (DOE) nécessaire à une estimation correcte du quantile. Cette approche permet de maitriser le coût de calcul tout en améliorant drastiquement les performances par rapport à l'approche d'estimation classique. Ainsi, dans le cadre du cas test final, nous avons identifié 4 fois plus de valeurs extrêmes de contraintes tout en réduisant d'un facteur 2.5 le nombre d'appel au modèle coûteux.

2. Pistes de réflexion

Nous avons considéré des modèles ayant une dimension, i.e. un nombre de variables d'entrée, relativement modeste. Nous soupçonnons la technique de SC associée au krigeage d'être en difficulté pour des problèmes en très grandes dimensions à cause du krigeage. Il n'existe pas de dimension minimum au-delà duquel le qualificatif de « grande dimension » est unanimement reconnu. Nous pouvons tout de même donner comme ordre de grandeur une vingtaine de variables. Un tel nombre de variables est en pratique rapidement atteignable, et mériterait d'être étudié.

La SC associée au krigeage apporte un gain de performance dans la recherche de valeurs extrêmes par rapport à l'utilisation directe krigeage seul. Il existe une autre approche, aussi basée sur un échantillonnage sélectif des entrées. Il s'agit de SMs dits adaptatifs. Le principe est de construire successivement plusieurs modèles de krigeage. Ces modèles de krigeage successifs se construisent sur un nouveau DOE enrichi par les entrées/sorties identifiées par le modèle de krigeage précédent comme valeur extrêmes. Le krigeage se prête d'autant plus facilement à ce processus adaptatif que, par construction, nous disposons d'une estimation d'erreur locale (variance du processus). Cette approche peut produire d'excellents résultats et être nettement supérieure, non seulement au krigeage seul, mais

potentiellement au krigeage associé à la SC. Cependant, la question du nombre de réalisations enrichissant chaque DOE reste ouverte. De plus, la solution adaptative souffrira encore plus sévèrement des mêmes problèmes en grande dimension que la version classique utilisée ici.

Dans ce chapitre nous avons supposé que nous disposions d'un modèle déterministe et surtout exact du phénomène. En électromagnétisme, la résolution numérique des équations de Maxwell produit effectivement des prédictions d'une très bonne précision. Cependant, une telle modélisation peut se révéler totalement irréaliste dans certaines situations. Par exemple, certains systèmes sont tout simplement trop grands et trop complexes pour être modélisés (fusées, avions, etc.), ne serait-ce que pour simuler une seule réalisation. De plus, des variables d'entrées dites « cachées », i.e. inconnues, peuvent exister et provoquer une erreur non négligeable du modèle. Une piste de recherche intéressante serait de se pencher sur le cas où le modèle exact ne l'est en réalité qu'à une certaine précision près, potentiellement inconnue.

CHAPITRE IV. ESTIMATION DE LA PROBABILITE DE DEFAILLANCE D'EQUIPEMENTS ELECTRONIQUES EN CAS D'AGRESSION ELECTROMAGNETIQUE

I. Introduction

Lorsqu'un équipement électronique est agressé par une interférence électromagnétique intentionnelle, une défaillance peut survenir. Malheureusement, non seulement les caractéristiques internes de l'équipement sont mal connues, mais celles de l'agression le sont aussi. Par conséquent, le risque de défaillance doit être quantifié en termes de probabilité. L'objectif est de calculer la probabilité que l'équipement subisse une défaillance en prenant en compte les degrés d'incertitude quant aux caractéristiques de l'équipement et de l'agression.

Littéralement, une défaillance résulte d'une contrainte supérieure au seuil de susceptibilité. Sachant les distributions de la contrainte (C) et de la susceptibilité (S), la probabilité de défaillance est donnée par l'intégrale suivante :

$$p_F = \mathbb{P}\{C \geq S\} = \int_{-\infty}^{\infty} \mathbb{P}\{C \geq x \cap S = x\}dx = \int_{-\infty}^{\infty} \big(1 - F_C(x)\big) f_S(x)dx. \tag{60}$$

Cette intégrale porte sur la fonction de répartition de la contrainte F_c et sur la densité de probabilité de la susceptibilité f_s. Dans le cadre d'une faible probabilité de défaillance (mais néanmoins non négligeable) les contributions majeures à cette intégrale proviennent conjointement de valeurs extrêmes supérieures de contrainte et extrêmes inférieures de susceptibilité.

Précédemment (Chapitre II), nous avons traité les incertitudes concernant l'équipement à l'aide d'une estimation de la distribution de probabilité de sa susceptibilité. Cette estimation provient de tests expérimentaux de susceptibilité en nombre limité, compte tenu de l'impossibilité à modéliser le processus de défaillance. La méthode que nous avons proposée repose sur un encadrement probabiliste de la distribution de susceptibilité. Ensuite (Chapitre III), nous avons traité des incertitudes concernant l'agression en estimant la distribution de probabilité des niveaux de contraintes extrêmes d'agression de l'équipement. L'approche que nous avons développée à cette fin repose sur l'articulation d'un SM (ici un krigeage) avec la méthode dite de stratification contrôlée pour accélérer la recherche de valeurs extrêmes. Les études de la susceptibilité et de la contrainte (Chapitres II et III) ont donc été traitées séparément l'une de l'autre. Cependant, la probabilité de défaillance résulte de la combinaison de ces deux distributions.

Ce chapitre rassemble dans un cas d'étude final, l'ensemble des approches proposées concernant l'étude de la contrainte d'une part, et de la susceptibilité d'autre part. Concernant l'étude de ces deux estimations, nous resterons succincts puisque l'objectif est ici l'exploitation de ces estimations afin d'obtenir la probabilité de défaillance finale. Le lecteur peut se référer aux chapitres correspondants, i.e. le chapitre II pour la susceptibilité et le chapitre III pour la contrainte. Si le principe des estimations est inchangé, quelques précautions doivent cependant être prises. En effet, la définition des niveaux de contrainte et de susceptibilité dépend de l'impédance d'entrée des équipements agressés. Afin d'éviter cette dépendance, nous allons apporter quelques précisions supplémentaires sur le processus de mesure des seuils de susceptibilité. Il ne s'agit pas de modifier le processus de mesure ni leur traitement mais d'apporter quelques précisions, omises au chapitre II dans un souci de simplification. Ce processus, est une forme de normalisation des seuils de susceptibilité par référence à une impédance fixe.

Ce chapitre est organisé en 3 sections, hormis l'introduction (section I) et la conclusion (section V). Tout d'abord (section II), nous présentons le cas d'étude final pour lequel nous souhaitons estimer la probabilité de défaillance. Puis (section III), nous détaillons la méthodologie utilisée pour obtenir une référence de contrainte /susceptibilité et surtout une valeur de référence de la probabilité

de défaillance. Enfin (section IV), nous présentons les résultats lorsque nous appliquons les méthodes proposées dans les deux précédents chapitres avec pour objectif l'estimation de la probabilité de défaillance.

II. Présentation du cas d'étude

1. Equipement soumis à une agression électromagnétique et chemin de couplage associé

Une « coque » (un parallélépipède rectangle) se situe dans un environnement réverbérant, i.e. une cavité électromagnétique comportant une ouverture circulaire sur sa face supérieure. Un cylindre se trouve au-dessus de l'une des faces de la coque. Le cylindre ainsi que les faces de la coque étant des conducteurs parfaits, il s'agit d'une antenne monopôle au-dessus d'un plan de masse. Une onde plane, d'incidence normale, illumine la cavité. Cette onde peut être considérée comme une agression de type IEMI, certes grossière, car harmonique. La cavité, l'antenne et la coque sont représentées à la Figure 60.

Nous supposons qu'un équipement se trouve à l'intérieur de la coque qui est un blindage parfait. La base de l'antenne est légèrement surélevée par rapport à la face considérée, et reliée à la coque par une charge. Nous supposons que cette charge relie cette antenne à un équipement sensible à l'intérieur de la coque. Cette charge représente l'impédance d'entrée normalisée de l'équipement. Le courant traversant cette charge équivaut à un courant à l'entrée de l'équipement qui risque de provoquer une défaillance. Une vue se focalisant sur l'antenne et la coque se trouve à la Figure 61.

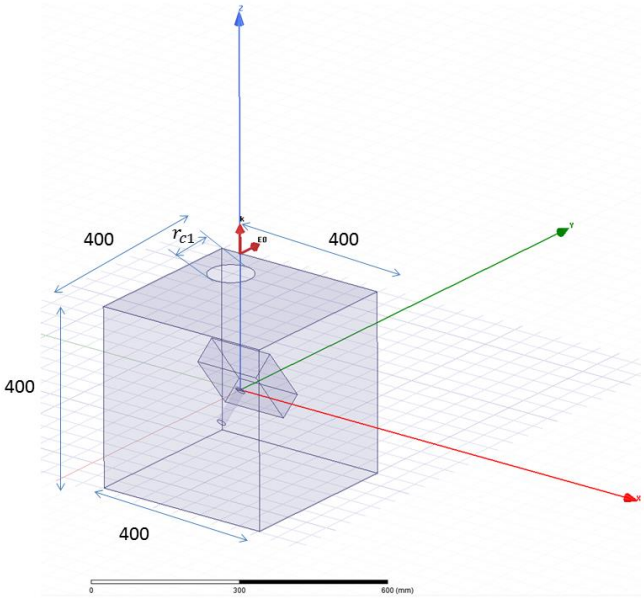

Figure 60 : Géométrie de la cavité percée d'une ouverture et de la coque surmontée par l'antenne.

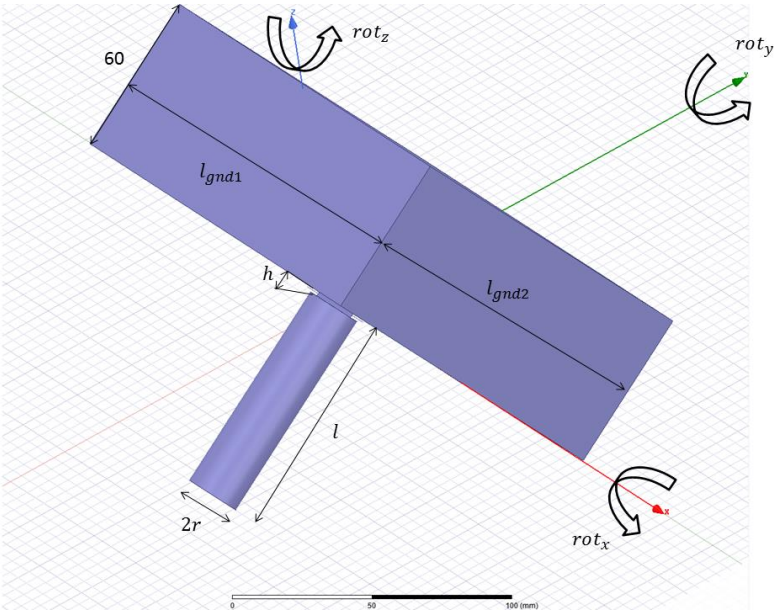

Figure 61 : Zoom de la Figure 60 sur l'antenne et la coque à l'intérieur de laquelle se trouve un équipement sensible (non figuré sur le schéma).

Concernant le choix de l'équipement à mettre à l'intérieur de la coque, trois possibilités s'offraient à nous :

1. Trouver une famille d'équipements pouvant être dans un environnement similaire. Cela implique d'effectuer une nouvelle campagne de mesures.
2. Définir un équipement virtuel afin d'effectuer la campagne de mesure en simulation.
3. Ne pas définir explicitement d'équipements mais réutiliser les mesures déjà effectuées dans le chapitre II.

La troisième solution, la plus simple, a été choisie. Les équipements à l'intérieur ne sont donc pas explicitement définis, et la distribution de susceptibilité sera celle des alimentations, déjà estimée dans le chapitre II. Une transformation de cette distribution sera tout de même appliquée pour être compatible au cas présent compte tenu de la présence de la charge normalisée à l'entrée de l'équipement. L'obtention d'une référence de la distribution de susceptibilité pour ce cas final est traitée en détail dans la section suivante de ce chapitre.

Le chemin de couplage de l'agression électromagnétique vers l'équipement est donc le suivant : suite à l'illumination par l'IEMI, une partie du champ EM pénètre à l'intérieur de la cavité par l'ouverture circulaire, pour finalement générer un courant sur l'antenne, qui est mesuré sur la charge. Les grandeurs caractérisant ce chemin sont reportées dans le Tableau 18. La plupart d'entre elles sont incertaines et donc modélisées par une variable aléatoire ayant une distribution uniforme ou gaussienne.

Tableau 18 : Entrées du modèle de l'antenne à l'intérieur de la cavité percée.

Variable	Description	Nominale	Distribution	Unité
f	fréquence	868	Fixe	*MHz*
E0	Amplitude de l'onde incidente	50	Fixe	*kV/m*
Z	Impédence de charge	50	Fixe	*Ω*
r	Rayon de l'antenne	10	$\mathcal{N}(10,5)$	*mm*
l	Longueur de l'antenne	81	$\mathcal{N}(81,40.5)$	
h	Distance entre la base de l'antenne et le plan de masse	2	$\mathcal{N}(2,1)$	
r_c1	Rayon de l'ouverture	50	$\mathcal{N}(50,25)$	
l_gnd1	Longueur de la coque	150	$U(140,160)$	
l_gnd2	Largeur de la coque	150		
rot_x	Rotation autour de l'axe x	0	$U(-180,180)$	*deg*
rot_y	Rotation autour de l'axe y	0		
rot_z	Rotation autour de l'axe z	0		

2. Mise en cohérence des niveaux de contrainte et de susceptibilité

Les variables aléatoires de la contrainte et de la susceptibilité ne sont pas strictement indépendantes, car elles dépendent toutes deux de la valeur de l'impédance d'entrée de l'équipement considéré. En effet, le caractère plus ou moins basse impédance de l'équipement (variable d'un équipement à l'autre), possède un impact sur le niveau de contrainte en courant ou en tension à l'entrée de l'équipement. Lorsque la contrainte et la susceptibilité sont traitées conjointement, il est nécessaire d'effectuer une normalisation au moyen d'une impédance de référence commune. Pour cela, nous normalisons les mesures de susceptibilité. Ceci permettra, lors de l'étude de la contrainte, de considérer l'impédance d'entrée de l'équipement fixe et égale à la valeur de normalisation choisie pour l'étude de la contrainte.

La procédure de normalisation des seuils de susceptibilité est schématisée dans la Figure 62 et se déroule en trois étapes. Cette procédure était en fait déjà utilisée dans le chapitre II. Les deux premières étapes n'avaient pas été citées pour plus de clarté.

1. Choisir un équipement (d'indice e) dont l'impédance d'entrée est Z_e. Augmenter E_e jusqu'à observer une susceptibilité. Au niveau de susceptibilité, relever le seuil S_e.
2. Remplacer l'équipement par Z_{ref} (e.g. 50 Ω), le réglage du générateur est inchangé par rapport à la fin de l'étape 1. Relever le seuil S_e^*.
3. Répéter les deux premières étapes sur plusieurs versions de l'équipement, afin de collecter des données de susceptibilité formant un ensemble de niveaux de susceptibilité mesurés.

Ainsi, les différents niveaux de susceptibilité *Se* ont été rendus indépendants de l'impédance d'entrée. Lors de la simulation de la contrainte nous allons pouvoir remplacer les équipements par une impédance fixe. Sa valeur sera identique à celle utilisée pour la normalisation (dans notre cas 50 Ω).

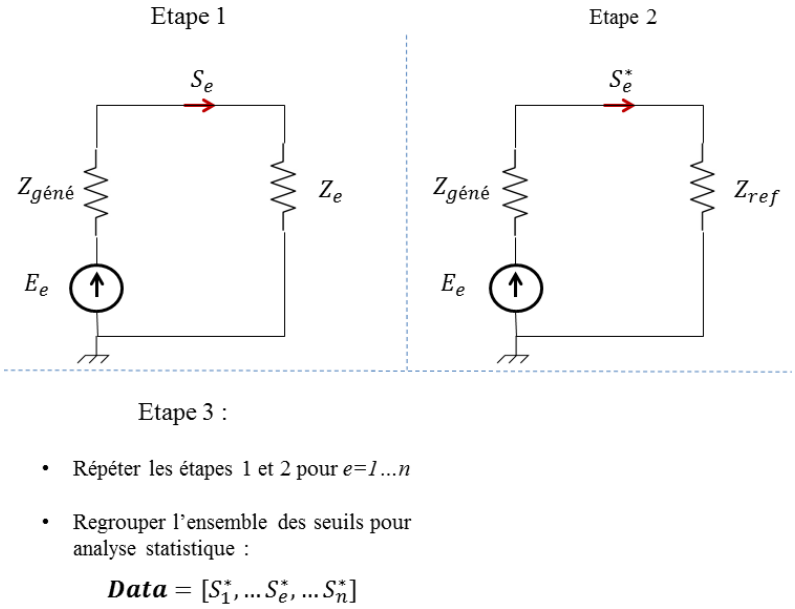

Figure 62 : Principe de normalisation des mesures de susceptibilité.

III. Obtention d'une probabilité de défaillance de référence

A présent, nous allons chercher à obtenir une référence pour la distribution de la contrainte et de la susceptibilité, afin d'obtenir une probabilité de défaillance de référence. Cette référence nous sera utile pour tester nos méthodes sur de petits échantillons. Evidemment, aucune référence n'est généralement disponible en situation réelle.

1. Contrainte de référence

Un grand échantillon de réalisations de sortie a été simulé à partir de 1000 réalisations d'entrées. Nous allons ensuite trouver la meilleure distribution qui s'ajuste à cet échantillon de référence. Un test statistique est donc effectué. L'échantillon est suffisamment grand pour obtenir des résultats fiables, quelle que soit la méthode d'inférence choisie ou le type de test d'ajustement.

Nous avons effectué le test statistique d'Anderson-Darling avec comme hypothèse nulle l'appartenance à une loi de distribution hypothétique. 25 lois répertoriées dans Matlab ont été successivement testées en tant qu'hypothèse nulle par la fonction Matlab *adtest*. La p-valeur sert ensuite de critère d'acceptation ou de rejet [141]. Rappelons qu'un test rejette l'hypothèse nulle lorsque la p-valeur est inférieure à une certaine probabilité de confiance (ici 5 %).

Nous traçons les p-valeurs sur la Figure 63 pour les 25 lois testées. Nous en concluons que seules 4 distributions ne sont pas clairement rejetées par le test (Gamma, Half-normal, Nakagami et Weibull). La p-valeur de certaines lois (Beta, Binomiale, Burr, Negative Binomiale, Multinomiale, Piece Wise Linear, Triangular) n'apparait pas car il est impossible d'effectuer le test pour des questions de support ou de continuité. Dans ces cas-là, il est évident que l'hypothèse nulle doit être rejetée.

Dans la Figure 64 nous traçons la *pdf* des lois non-rejetées. Nous retenons la distribution de Weibull comme étant la distribution de référence de la contrainte avec pour paramètres $\gamma = 1.0057$ et $\beta = 1.3445$. En effet, sa p-valeur et donc sa vraisemblance sous hypothèse nulle, sont importantes. La distribution Nakagami [142][143][144] aurait aussi pu convenir. La loi Gamma n'est acceptée qu'à quelques pourcents près (p-valeur à peine au-dessus de 5%) ce qui est insatisfaisant.

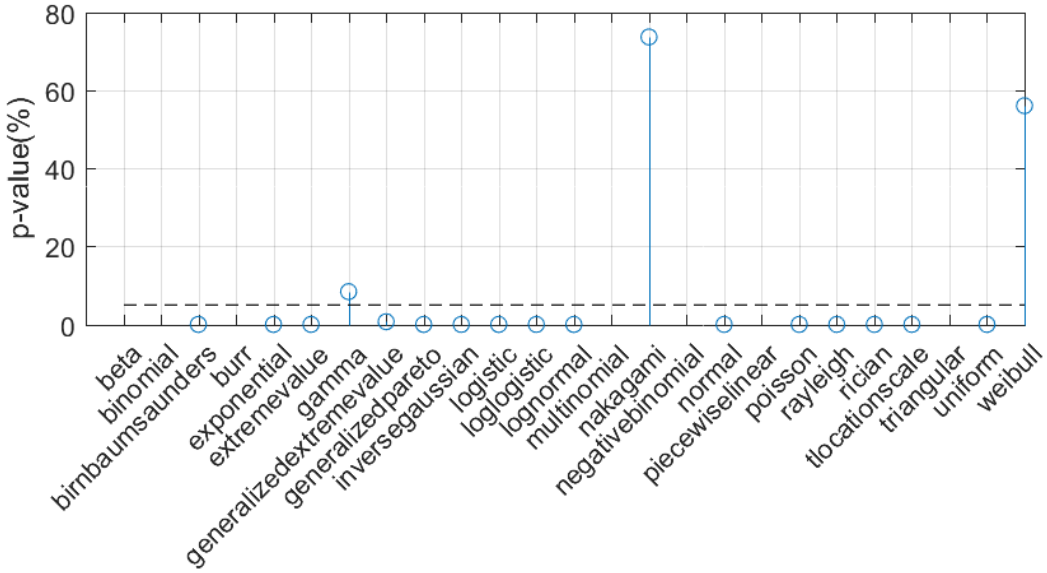

Figure 63 : p-valeurs du test statistique AD (IMV) à partir de l'échantillon de contrainte de référence pour plusieurs hypothèses de distribution.

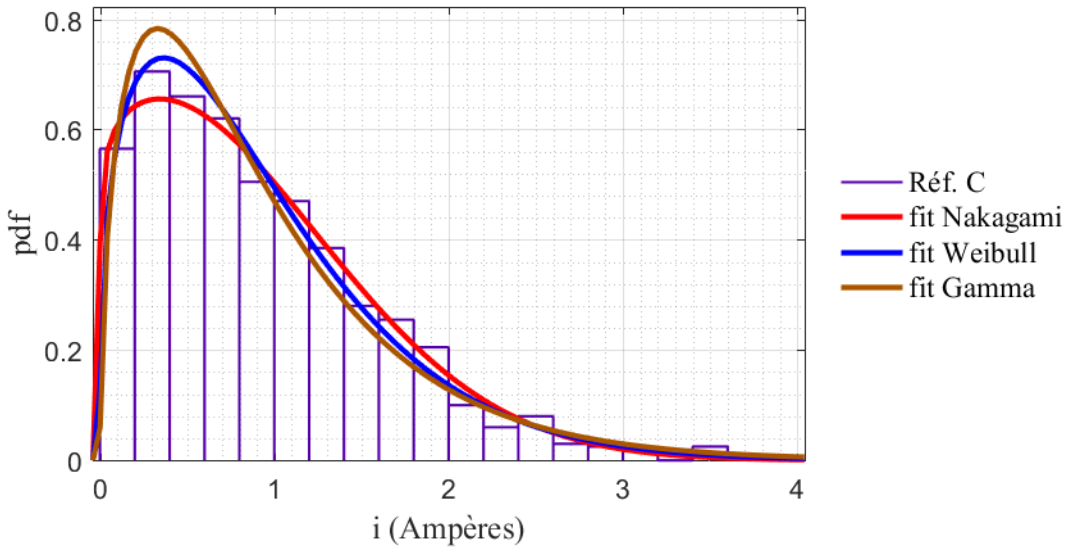

Figure 64 : Histogramme de l'échantillon de référence de la contrainte (C) en Ampères, ainsi que les meilleurs ajustements pour chacune des lois non rejetées par le test d'AD (IMV).

2. Susceptibilité de référence

La distribution de susceptibilité théorique s'inspire de la distribution de susceptibilité estimée dans le chapitre II. Il s'agissait de la mesure de susceptibilité d'alimentations. La distribution estimée était une loi Log-normale, de paramètres $a = 2.06$ et $b = 0.24$. Nous allons considérer ici qu'il ne s'agit pas de paramètres estimés, mais de valeurs de paramètres exactes.

3. Probabilité de défaillance de référence

Connaissant la distribution de la contrainte de référence donnée par une loi de Weibull de paramètre ($\gamma = 1.0057$, $\beta = 1.3445$) et la distribution de la susceptibilité de référence donnée par une loi Log-normale de paramètres ($a = 2.06$, $b = 0.24$), nous obtenons avec (60) une probabilité de défaillance de référence égale à environ 0.138 %.

Dans la Figure 65, nous traçons les distributions de référence, celle de la contrainte ainsi que celle de la susceptibilité. De plus, nous traçons la *pdf* de la défaillance *D* associée un niveau de courant *i* définie par $C \geq S$ tel que $P(D = i) = P(\, C = i | C \geq S)/P(C \geq S)$.

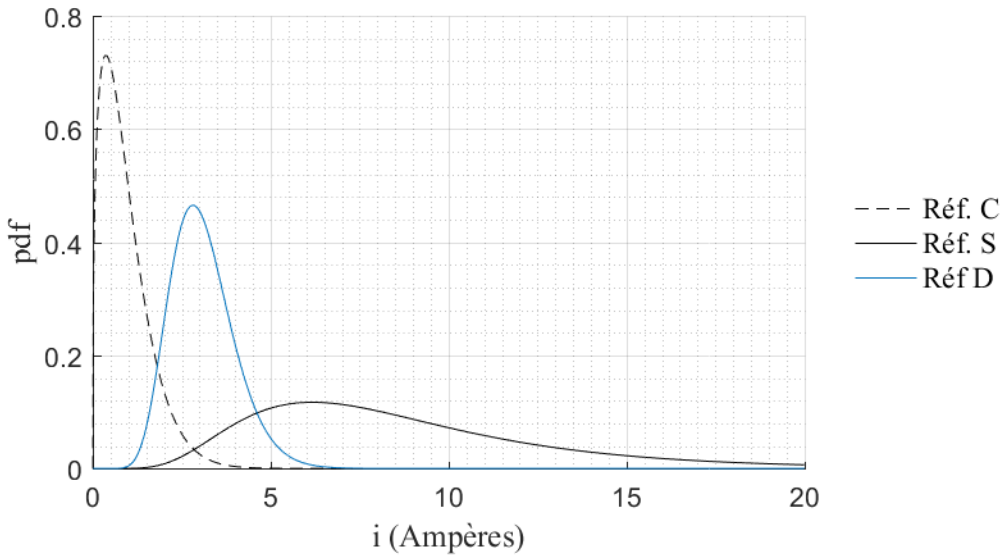

Figure 65 : Densités de probabilité de la contrainte (C) et de la susceptibilité (S) de référence (Réf.) en Ampère et densité de probabilité de la défaillance (D) qui en découle.

IV. Estimation de la probabilité de défaillance

Estimons à présent la distribution de la susceptibilité d'une part, puis celle de la contrainte d'autre part, à partir de petits échantillons. L'objectif est d'estimer la probabilité de défaillance dont nous avons une référence, calculée précédemment.

Concernant l'étude de la susceptibilité, aucun changement majeur n'est apporté par rapport au Chapitre II. En effet, un petit échantillon de taille n_S est issu d'un tirage aléatoire dans la distribution de susceptibilité de référence qui est une loi log-normale de paramètres a=2.06 et b= 0.24. Puis l'inférence statistique est appliquée afin d'obtenir deux distributions encadrant la distribution de susceptibilité.

Pour l'étude de la contrainte, nous procédons à un tirage aléatoire de n_K réalisations parmi l'échantillon de référence afin de construire un krigeage. Il s'agit donc d'un DOE échantillonné à partir de l'échantillon de référence qui contient 1000 réalisations. Enfin, la SC est appliquée, ce qui nécessite la simulation par HFSS de n_{SC} réalisations supplémentaires, avec n_{SC}=n_K. Ainsi, un budget de simulation de n_C=n_K+n_{SC} est utilisé pour estimer la contrainte. Par simplicité, nous n'allons pas utiliser l'algorithme proposé au chapitre III section VI qui comportait une stratégie d'enrichissement, à la fois de la partie krigeage et de la partie SC. Cette suppression se justifie par le fait que l'objectif de ce chapitre est le traitement des estimations des contraintes, non leur estimation. Malheureusement, cela implique de se contenter d'un choix arbitraire concernant la taille du DOE pour construire le krigeage.

Dans les études précédentes, la K-SC était utilisée afin de répondre à l'objectif qui était d'identifier un plus grand nombre de valeurs extrêmes afin d'estimer un quantile dont la probabilité était connue a priori. Les quantiles définissant les limites des strates étaient donc connus a priori. Cependant, l'objectif est ici plus précis : il s'agit de cibler plus particulièrement les valeurs extrêmes qui ont un impact sur le calcul de la probabilité de défaillance. Ceci est possible en exploitant les distributions empiriques de contrainte et de susceptibilité. Nous présentons donc une nouvelle stratégie de stratification adaptée au problème posé.

1. Nouvelle stratégie de stratification pour la SC

L'allocation selon les strates est toujours uniforme mais les probabilités délimitant les strates sont maintenant adaptées au nouvel objectif. En effet, l'estimation de la distribution de la contrainte nécessite d'être précise uniquement pour les réalisations pouvant provoquer une défaillance. C'est pourquoi nous optons pour une stratégie de stratification permettant d'échantillonner les nouvelles réalisations extrêmes qui auront le plus d'impact sur l'estimation de la probabilité de défaillance. Cette stratégie repose sur un choix judicieux des probabilités définissant les bornes des strates.

Grâce au théorème de Bayes, puisque C et S sont indépendantes, nous pouvons écrire que :

$$\mathbb{P}\{D = i\} = \mathbb{P}\{C = i | C \geq S\} = \frac{\mathbb{P}\{C=i\}\mathbb{P}\{S \leq i\}}{\mathbb{P}\{C \geq S\}}. \tag{61}$$

Nous possédons une estimation de la *pdf* de C grâce au krigeage. De plus, l'inférence fournit deux *cdf*, encadrant les seuils de S, correspondant à un pire cas et à un cas optimiste. Le pire cas correspond à la *cdf* qui implique une probabilité de défaillance plus élevée. Il s'agit donc de la *cdf* la plus proche de celle de la contrainte. Disposant d'une estimation de la distribution empirique de C (i.e. $\hat{f}_C$) et d'une estimation de la *cdf* de S (i.e. $\hat{F}_S$), la distribution de probabilité de D peut être estimée ainsi :

$$\hat{f}_D(i) = \frac{\hat{f}_C(i)\hat{F}_S(i)}{\int \hat{f}_C(i)\hat{F}_S(i)di}. \tag{62}$$

La *cdf* de D peut alors être déduite à partir de la *pdf*, et donc les quartiles de D (i.e. les quantiles 0.25, 0.5 et 0.75) regroupés dans le vecteur $\widehat{\boldsymbol{d}}$. Le vecteur $\boldsymbol{A}$ définissant les limites des bornes pour la SC est donc :

$$\boldsymbol{A} = [0, \hat{F}_C^{-1}(\widehat{\boldsymbol{d}}), 1]. \tag{63}$$

Le budget de simulation de la SC (i.e. n_{SC}) est réparti équitablement entre les 4 strates. Ainsi, N_j réalisations sont allouées par strate :

$$N_j = \left\lfloor \frac{n_{SC}}{4} \right\rfloor + mod(n_{SC}, 4)\mathbb{I}_{j=4}, \quad j = 1,..,4. \tag{64}$$

2. Résultats de l'estimation de la probabilité de défaillance

Nous présentons à présent la probabilité de défaillance estimée pour quelques valeurs de taille d'échantillons consacrés à l'estimation de la contrainte (n_C) et de la susceptibilité (n_S). Pour chaque couple (n_C,n_S), la probabilité de défaillance est calculée à partir des deux distributions encadrant la distribution de susceptibilité. Nous obtenons ainsi des bornes encadrant la probabilité de défaillance. Comme tout estimateur, celui des bornes présente nécessairement une fluctuation aléatoire. Cette fluctuation n'a cependant pas été quantifiée, faute de temps, mais pourrait l'être en pratique à partir de l'exploitation du krigeage.

Le Tableau 19 synthétise l'ensemble des résultats pour des valeurs de n_s=10,50,100 et de n_c=100,500,1000. Quelques observations intéressantes peuvent être formulées.

A partir de n_c=500, la probabilité de défaillance de référence est bien située dans l'intervalle des valeurs de probabilité de défaillance estimée. En deçà, pour nc=100, la borne inférieure peut-être largement sous-estimée et la borne supérieure également (pour les cas n_s=50 et n_s=100). Par conséquent, bien que la loi de susceptibilité soit bien estimée, particulièrement lorsque n_S=100, l'estimation de la probabilité de défaillance reste sous-estimée par rapport à la valeur de référence. Ceci tend à montrer que la taille de l'échantillon est trop faible pour qu'un nombre significatif de valeurs extrêmes d'intérêt soit identifié, malgré l'utilisation de la stratification contrôlée.

D'après 100 échantillons de taille n_K=500 et 100 de taille n_s=100, les probabilités ***A*** des quantiles à cibler sont, en moyenne, égales à [0 , 0.94787 , 0.96977 , 0.97742 , 1]. La taille minimale du DOE pour construire un SM se déduit de la probabilité du quantile à estimer (méthodologie de la loi binomiale négative présentée au chapitre III). D'après cette méthodologie, l'estimation des quantiles de probabilités 0.94787 , 0.96977 et 0.97742 nécessite une taille minimale de DOE de 56, 98, et 132 respectivement. Ceci explique les mauvaises performances dans le Tableau 19 lorsque n_c=100. Dans ce cas, un DOE de taille n_K =50 a été utilisé pour construire le krigeage, ce qui n'est même pas suffisant pour estimer correctement le quantile de probabilité 0.94787.

Nous remarquons également que la précision de l'encadrement de la probabilité de défaillance dépend fortement de l'estimation de la susceptibilité. Ainsi, lorsque n_S augmente, à n_C identique, la précision de cette estimation augmente très significativement. Ceci est évidemment lié à l'impossibilité d'inférer une statistique précise à partir d'un échantillon de faible taille (n_s=10).

Tableau 19 : Probabilité de défaillance estimée (%), la référence étant de 0.13795 %, en fonction du nombre de réalisation utilisées pour estimer la contrainte (n_C) et celui pour la susceptibilité (n_S).

n_S \ n_C	100	500	1000
10	[0.00032094 , 1.8674]	[0.052384 , 4.3463]	[0.036087 , 3.1713]
50	[0.00045852 , 0.068394]	[0.032288 , 0.51603]	[0.050134 , 0.61374]
100	[0.0077149 , 0.11868]	[0.046715 , 0.26505]	[0.035661 , 0.21623]

V. Conclusion et piste de réflexion

Dans ce dernier chapitre applicatif, nous avons montré comment valoriser les méthodes traitées dans cette thèse afin d'estimer la probabilité de défaillance en cas d'agression électromagnétique intentionnelle. In fine, c'est l'information la plus importante pour une analyse du risque de défaillance face à une IEMI.

L'encadrement de l'estimation de la probabilité de défaillance est principalement lié à la dimension de l'échantillon de susceptibilité qui détermine la précision de l'encadrement de la loi de susceptibilité. La taille de l'échantillon de contrainte, au-delà d'une taille minimale permettant de reproduire correctement la distribution de valeurs extrêmes (en fonction du quantile recherché et donc du niveau de probabilité recherché) ne joue pas un rôle prédominant dans la qualité de l'estimation pour ce cas test.

La taille minimale du DOE pourrait être choisie par une stratégie d'enrichissement du DOE. Une telle stratégie peut s'inspirer de celle du chapitre III, en l'adaptant à la recherche d'événements provoquant la défaillance, e.g. en utilisant la stratégie de stratification présentée dans ce chapitre.

CHAPITRE V. CONCLUSION GENERALE

I. Rappel de l'objectif et de la stratégie

L'objectif de cette thèse était d'évaluer le risque de défaillance d'un équipement électronique en cas d'agression électromagnétique intentionnelle. Deux difficultés majeures rendent cette estimation complexe. D'une part, les incertitudes épistémiques associées à la description du modèle de la contrainte n'autorisent qu'une estimation probabiliste de la contrainte qui résulte de l'agression. D'autre part, il est illusoire, encore à ce jour, de modéliser le fonctionnement d'un équipement pour en évaluer la susceptibilité. Seules des mesures permettent d'évaluer les fluctuations de niveaux de susceptibilité associées aux incertitudes de fabrication de l'équipement.

L'approche adoptée dans cette thèse consiste à évaluer la probabilité de défaillance à partir d'une inférence statistique à partir de mesures de susceptibilité et de modèles de contrainte aux paramètres incertains. La première partie de nos travaux (Chapitre II) a été consacrée à l'étude de la susceptibilité qui nous a permis d'aboutir à un encadrement vraisemblable de la densité de la loi de susceptibilité. Le second volet (Chapitre III) nous a permis d'examiner la contrainte sous un angle également probabiliste. En présence de paramètres incertains, il s'agit d'évaluer le plus efficacement possible, la distribution de la contrainte et plus particulièrement la distribution des valeurs extrêmes de celle-ci. Enfin, la combinaison des approches développées dans ces deux chapitres a permis de déduire la probabilité de défaillance pour un cas test (Chapitre IV). Cette conclusion générale synthétise les résultats obtenus sur ces trois volets de la thèse et dresse pour chacun d'entre eux des perspectives possibles.

II. Synthèse de l'étude de la susceptibilité

Le chapitre II a été consacré à l'inférence statistique de la distribution de susceptibilité d'équipements électroniques à partir d'un échantillon très limité de seuils de susceptibilité mesurés. En effet, le nombre d'équipements disponibles pour les essais est généralement faible à cause des coûts des campagnes de mesures.

Nous avons comparé deux types d'inférence paramétrique : bayésienne (IB) et celle du maximum de vraisemblance (IMV, méthode dite « fréquentiste »), bien que cette dernière puisse être traitée comme un cas particulier de la première. Les quatre études menées dans ce chapitre suggèrent qu'à partir d'environ 40 équipements, la différence entre les deux méthodes d'inférence n'est pas significative. De plus, il n'est pas nécessairement rentable d'augmenter la taille d'échantillon au-delà de 40 équipements environ, car une augmentation beaucoup plus importante est nécessaire afin d'obtenir une précision significativement plus intéressante. Pour des échantillons de plus petite taille, il n'est pas possible d'identifier clairement la méthode à privilégier. Au-delà du choix de la méthode d'inférence, il est important de ne pas se contenter d'une estimation la plus probable au risque de négliger la grande incertitude due à la petite taille de l'échantillon. A ce titre, l'IMV tout comme l'IB fournit un intervalle de confiance des paramètres de la distribution. Une approche pertinente pour l'analyse du risque consiste donc à utiliser ces intervalles pour encadrer la distribution de susceptibilité, que l'estimation provienne d'une IB ou d'une IMV. Cette approche permet de rendre compte convenablement de la vraisemblance de la distribution de susceptibilité à partir d'un faible nombre de données.

Une perspective intéressante de ces travaux serait de s'affranchir de l'obligation de choisir une hypothèse de loi support (Normale, Weibull, etc.), inconvénient majeur des inférences paramétriques. La méthode de Bootstrap, qui ne nécessite aucune hypothèse de ce type, pourrait constituer une alternative intéressante aux inférences paramétriques, puisqu'elle élimine de fait le risque d'erreur du choix a priori de la loi support.

III. Synthèse de l'étude de la contrainte

Le chapitre III visait à estimer la contrainte en cas d'IEMI en se focalisant sur l'estimation des valeurs extrêmes. L'objectif était d'améliorer la technique standard de MCS appliquée au modèle exact (méthode empirique simple et robuste mais à convergence lente). La stratification contrôlée, méthode déjà identifiée dans de précédents travaux, est une technique spécifiquement conçue dans ce but.

Cependant, la stratification contrôlée nécessite un modèle approximant le modèle de référence souvent coûteux, i.e. un métamodèle, afin d'assurer un contrôle efficace de la stratification. Aucune approche générale n'est proposée dans la littérature à ce sujet. C'est pourquoi, nous avons cherché un SM générique à même de remplir cet objectif selon un critère de corrélation propre à la stratification contrôlée. Après comparaison de quelques SMs les plus courants, nous avons établi que le krigeage était le plus pertinent et potentiellement efficace. Nous avons ensuite cherché dans quelle mesure la stratification contrôlée apportait un gain de performance par rapport au krigeage utilisé seul. Nous avons montré au travers de deux modèles que plus la non-linéarité du modèle est importante, plus la stratification contrôlée est pertinente. Il en va de même si les valeurs de sortie visées sont de plus en plus rares. Enfin, nous nous sommes intéressés à une implémentation pratique de la stratification contrôlée associée au krigeage, méthode que nous avons dénommée K-SC. Nous avons proposé un algorithme afin de maîtriser le coût de calcul tout en améliorant drastiquement les performances par rapport à l'approche standard en CEM. Cet algorithme permet, à partir d'une taille de DOE initiale fixée selon le quantile recherché, d'ajuster automatiquement celle-ci par enrichissement successif si nécessaire. Ainsi, dans le cadre du cas test étudié, nous avons identifié 4 fois plus de valeurs extrêmes de contrainte tout en réduisant d'un facteur 2.5 le nombre d'appel au modèle coûteux.

Une perspective possible à donner à ces travaux concerne les problèmes à grande dimension. Dans cette thèse, les modèles étudiés étaient tous sans exception de dimensions relativement peu élevées. En effet, le nombre de variables aléatoires d'entrée était toujours inférieur à une vingtaine. Or, dans de nombreux problèmes industriels, ce nombre peut facilement être largement supérieur (e.g. une centaine). Dans ces situations, construire un SM ordinaire, comme le krigeage, n'est plus négligeable en temps de calcul. Cependant, des méthodes dédiées comme la PLS ou une analyse de sensibilité peuvent permettre d'accélérer la construction. Au sein de la méthode K-SC, une perspective intéressante serait donc de modifier le krigeage par un krigeage couplé à la PLS, ce qui donnerait une nouvelle méthode (e.g. K-PLS-SC) dédiée à la recherche d'événements extrêmes en grandes dimensions.

IV. Synthèse de l'étude de la probabilité de défaillance

Dans le chapitre IV, nous avons montré comment valoriser les méthodes traitées dans cette thèse afin d'estimer la probabilité de défaillance en cas d'agression électromagnétique intentionnelle. In fine, c'est l'information la plus importante pour une analyse de risque.

Ce court chapitre a permis d'établir les performances de l'estimation de la probabilité de défaillance. Compte tenu de l'encadrement de la distribution de susceptibilité, la probabilité de défaillance est également encadrée. Nous montrons que l'intervalle de prédiction est crédible (au sens où il encadre la valeur de référence, évaluée pour le cas testé ici mais qui sera inconnue en pratique) si la taille de l'échantillon de calcul de la contrainte atteint une taille suffisante. La précision de l'intervalle, i.e. sa largeur, dépend ensuite de l'estimation de la susceptibilité.

Une prolongation naturelle de cette analyse serait de sélectionner a priori la taille minimale du DOE par une stratégie d'enrichissement successive d'un DOE initial choisi selon le quantile recherché. Une telle stratégie peut s'inspirer de celle du chapitre III, en l'adaptant à la recherche d'événements provoquant la défaillance, e.g. en utilisant la stratégie de stratification présentée dans le chapitre IV.

RÉFÉRENCES

[1] W. Radasky and E. Savage, "RADASKY, William et SAVAGE, Edward. Intentional electromagnetic interference (IEMI) and its impact on the US power grid," *Metatech*, 2010.

[2] P. Angskog, P. Nasman, and L.-G. Mattsson, "Resilience to Intentional Electromagnetic Interference Is Required for Connected Autonomous Vehicles," *IEEE Trans. Electromagn. Compat.*, pp. 1–8, 2018.

[3] R. R. Gaynutdinov and S. F. Chermoshentsev, "Study immunity to disturbance of electronic system aircraft by influences of intentional ultrashort electromagnetic pulses," in *2016 International Conference on Actual Problems of Electron Devices Engineering (APEDE)*, Saratov, 2016, pp. 1–6.

[4] M. Heddebaut, V. Deniau, J. Rioult, and C. Gransart, "Mitigation Techniques to Reduce the Vulnerability of Railway Signaling to Radiated Intentional EMI Emitted From a Train," *IEEE Trans. Electromagn. Compat.*, vol. 59, no. 3, pp. 845–852, Jun. 2017.

[5] H. Rausch, "Jamming Commercial Satellite Communications During Wartime: An Empirical Study," in *Fourth IEEE International Workshop on Information Assurance (IWIA'06)*, London, UK, 2006, pp. 109–118.

[6] M. Ivezic, "Intentional Electromagnetic Interference (IEMI) – the overlooked threat to IoT," *CSO*, 2018.

[7] F. Sabath, "What can be learned from documented Intentional Electromagnetic Interference (IEMI) attacks?," in *2011 XXXth URSI General Assembly and Scientific Symposium*, Istanbul, 2011, pp. 1–4.

[8] M. U. Karlssonx, F. Olsson, D. Aberg, and M. Jansson, "Bofors HPM blackout - a versatile and mobile L-band high power microwave system," in *2009 IEEE Pulsed Power Conference*, Washington, DC, 2009, pp. 499–501.

[9] D. Martin, "The Pentagon's Ray Gun," *cbs news*, 2008.

[10] "Arme à énergie dirigée," *Wikipedia*.

[11] J. S. Foster *et al.*, "Report of the Commission to Assess the Threat to the United States from Electromagnetic Pulse (EMP) Attack," 2004.

[12] "Israël songerait à lancer une impulsion électromagnétique sur l'Iran," *Le Monde*, 2012.

[13] R. P. Feynman *et al.*, *Le cours de physique de Feynman*. 2017.

[14] M. Sallak, F. Aguirre, and W. Schon, "Incertitudes aléatoires et épistémiques, comment les distinguer et les manipuler dans les études de fiabilité ?," presented at the QUALITA2013, Compiègne, 2013.

[15] C. F. M. Carobbi, S. Lallechere, and L. R. Arnaut, "Review of Uncertainty Quantification of Measurement and Computational Modeling in EMC Part I: Measurement Uncertainty," *IEEE Trans. Electromagn. Compat.*, pp. 1–9, 2019.

[16] P. Beeckman, "Uncertainty about uncertainties along the EMC-compliance chain," Barcelona, 2019.

[17] "Evaluation of measurement data — Guide to the expression of uncertainty in measurement," Working Group 1 of the Joint Committee for Guides in Metrology, 2008.

[18] *CISPR TR 16-4-1:2009 Specification for radio disturbance and immunity measuring apparatus and methods - Part 4-1: Uncertainties, statistics and limit modelling - Uncertainties in standardized EMC tests*.

[19] J. J. Goedbloed and P. A. Beeckman, "Uncertainty Analysis of the CISPR/A Radiated Emission Round-Robin Test Results," *IEEE Trans. Electromagn. Compat.*, vol. 46, no. 2, pp. 246–262, May 2004.

[20] S. Lalléchère, C. F. M. Carobbi, and L. R. Arnaut, "Review of Uncertainty Quantification of Measurement and Computational Modeling in EMC Part II: Computational Uncertainty," *IEEE Trans. Electromagn. Compat.*, pp. 1–8, 2019.

[21] M. Larbi, "Méthodes statistiques pour le calcul d'interférences électromagnétiques extrêmes au sein de systèmes complexes," PhD thesis, Rennes, INSA, 2016.

[22] D. Nitsch, M. Camp, F. Sabath, J.-L. ter Haseborg, and H. Garbe, "Susceptibility of Some Electronic Equipment to HPEM Threats," *IEEE Trans. Electromagn. Compat.*, vol. 46, no. 3, pp. 380–389, Aug. 2004.

[23] G. Spadacini and S. A. Pignari, "Numerical Assessment of Radiated Susceptibility of Twisted-Wire Pairs With Random Nonuniform Twisting," *IEEE Trans. Electromagn. Compat.*, vol. 55, no. 5, pp. 956–964, Oct.

2013.

[24] M. Mehri, N. Masoumi, and J. Rashed-Mohassel, “Trace Orientation Function for Statistical Prediction of PCB Radiated Susceptibility and Emission,” *IEEE Trans. Electromagn. Compat.*, vol. 57, no. 5, pp. 1168–1178, Oct. 2015.

[25] M. Mehri and N. Masoumi, “Statistical Prediction and Quantification of Radiated Susceptibility for Electronic Systems PCB in Electromagnetic Polluted Environments,” *IEEE Trans. Electromagn. Compat.*, vol. 59, no. 2, pp. 498–508, Apr. 2017.

[26] E. Genender, H. Garbe, and F. Sabath, “Probabilistic Risk Analysis Technique of Intentional Electromagnetic Interference at System Level,” *IEEE Trans. Electromagn. Compat.*, vol. 56, no. 1, pp. 200–207, Feb. 2014.

[27] C. F. M. Carobbi, “Bayesian Inference in Action in EMC—Fundamentals and Applications,” *IEEE Trans. Electromagn. Compat.*, vol. 59, no. 4, pp. 1114–1124, Aug. 2017.

[28] Chen Yuhao, Li Kejie, and Xie Yanzhao, “Bayesian assessment method of device-level electromagnetic pulse effect based on Markov Chain Monte Carlo,” in *2016 Asia-Pacific International Symposium on Electromagnetic Compatibility (APEMC)*, Shenzhen, China, 2016, pp. 659–661.

[29] L. M. Leemis and J. T. McQueston, “Univariate Distribution Relationships,” *Am. Stat.*, vol. 62, no. 1, pp. 45–53, Feb. 2008.

[30] E. T. Jaynes, “Prior Probabilities,” *IEEE Trans Syst. Sci. Cybern.*, vol. 4, no. 3, pp. 227–241, 1968.

[31] A. R. Syversveen, “Non informative Bayesian Priors. Interpretation And Problems With Construction And Applications,” *Prepr. Stat.*, vol. 3, pp. 1–11, 1998.

[32] A. Zellner, “Maximal data information prior distributions,” *New Dev. Appl. Bayesian Methods*, pp. 211–232, 1977.

[33] A. Jeffrey, “An invariant form for the prior probability in estimation problems,” presented at the Proceedings of the Royal Society of London a: mathematical, physical and engineering sciences, 1946, pp. 453–461.

[34] J. Berger and J. Bernardo, “On the development of the reference prior method,” *Bayesian Stat.*, vol. 4, no. 4, pp. 35–60, 1992.

[35] R. Yang and J. O. Berger, *A Catalog of Non informative Priors*, Institute of Statistics and Decision Sciences. Duke University, 1998.

[36] E. Fabrizi and C. Trivisano, “Bayesian Estimation of Log-Normal Means with Finite Quadratic Expected Loss,” *Bayesian Anal.*, vol. 7, no. 4, pp. 975–996, 2012.

[37] A. Zellner, “Bayesian and non-Bayesian analysis of the log-normal distribution and log-normal regression,” *J. Am. Stat. Assoc.*, vol. 66, no. 334, pp. 327–330, 1971.

[38] G. Casella and R. L. Berger, *Statistical Inference*, 2nd ed. .

[39] J.J. Ruch, “STATISTIQUE : ESTIMATION,” 2013.

[40] L. Perreault and B. Bobée, “Loi Weibull à deux paramètres Propriétés mathématiques et statistiques Estimation des paramètres et des quantiles Xr de période de retour T,” Scientifique 351, 1992.

[41] S. H. Zanakis, “Extended Pattern Search with Transformations for the Three-Parameter Weibull MLE Problem,” *Manag. Sci.*, vol. 25, no. 11, pp. 1149–1161, Nov. 1979.

[42] H. Hirose, “Bias Correction for the Maximum Likelihood Estimates in the Two-parameter Weibull Distribution,” *IEEE Trans Dielectr. Electr. Insul.*, vol. 6, no. 1, pp. 66–68, 1999.

[43] K. P. Burnham and D. R. Anderson, “Multimodel Inference: Understanding AIC and BIC in Model Selection,” *Sociol. Methods Res.*, vol. 33, no. 2, pp. 261–304, Nov. 2004.

[44] W. J. Conover, *Practical nonparametric statistics*, 3rd ed. New York Chichester: John Wiley, 1999.

[45] F. J. Massey, “The Kolmogorov-Smirnov Test for Goodness of Fit,” *J. Am. Stat. Assoc.*, vol. 46, no. 253, pp. 68–78, 1951.

[46] T. W. Anderson and D. A. Darling, “A Test of Goodness-of-Fit,” *J. Am. Stat. Assoc.*, vol. 49, no. 268, pp. 765–769, 1954.

[47] H. W. Lilliefors, “On the Kolmogorov-Smirnov Test for Normality with Mean and Variance Unknown,” *J. Am. Stat. Assoc.*, vol. 62, no. 318, p. 399, Jun. 1967.

[48] S. S. Shapiro and M. B. Wilk, “An Analysis of Variance Test for Normality (Complete Samples),” *Biometrika*, vol. 52, no. 3/4, p. 591, Dec. 1965.

[49] P. Royston, “Approximating the Shapiro-Wilk W-test for non-normality,” *Stat. Comput.*, vol. 2, no. 3, pp. 117–119, Sep. 1992.

[50] N. M. Razali and Y. B. Wah, “Power comparison of shapiro-wilk kolmogorov-smirnov lilliefors and anderson-darling tests,” *J. Stat. Model. Anal.*, vol. 2, no. 1, pp. 21–33, 2011.

[51] A. Pekgör, M. Erişoğlu, A. Karakoca, and Ü. Erişoğlu, “Empirical Type 1 Error Rate and Power Comparisons of Normality Tests with R,” *Cumhur. Sci. J.*, Sep. 2018.

[52] C. Pouant, “Caractérisation de la susceptibilité électromagnétique des étages d’entrée de composants électroniques,” Université Montpellier, 2015.

[53] E. L. Kaplan and P. Meier, “Nonparametric Estimation from Incomplete Observations,” *J. Am. Stat. Assoc.*, vol. 53, no. 282, p. 457, Jun. 1958.

[54] E. S. Banjanovic and J. W. Osborne, “Confidence Intervals for Effect Sizes: Applying Bootstrap Resampling,” *Pract. Assess. Res. Eval.*

[55] L. Labarbe, J. P. Adam, and J. M. Lopez, “Uncertainties Evaluation of Numerical Simulations with a FD-TD Solver – GORF3D,” presented at the ASIAEM, Xi’an, China, 2019.

[56] J. P. Adam and S. Vauchamp, “EMC on Large Aircrafts: Computer Simulations to Assist a Full Scale Experiment,” presented at the AMEREM, Santa Barbara, USA, 2018.

[57] J. P. C. Kleijnen, “Design Of Experiments: Overview,” in *2008 Winter Simulation Conference*, Miami, FL, USA, 2008, pp. 479–488.

[58] M. Stein, “Large Sample Properties of Simulations Using Latin Hypercube Sampling,” *Technometrics*, vol. 29, no. 2, p. 143, May 1987.

[59] B. Sudret, S. Marelli, and J. Wiart, “Surrogate models for uncertainty quantification: An overview,” in *2017 11th European Conference on Antennas and Propagation (EUCAP)*, Paris, France, 2017, pp. 793–797.

[60] A. Ajayi, C. Christopoulos, P. Sewell, G. A. Borges, and L. R. A. X. de Menezes, “Efficient computation of stochastic electromagnetic problems using unscented transforms,” *IET Sci. Meas. Technol.*, vol. 2, no. 2, pp. 88–95, Mar. 2008.

[61] C. Chauvière, J. S. Hesthaven, and L. Lurati, “Computational Modeling of Uncertainty in Time-Domain Electromagnetics,” *SIAM J. Sci. Comput.*, vol. 28, no. 2, pp. 751–775, Jan. 2006.

[62] P. Bonnet, F. Diouf, C. Chauvière, S. Lalléchère, M. Fogli, and F. Paladian, “Numerical simulation of a Reverberation Chamber with a stochastic collocation method,” *Comptes Rendus Phys.*, vol. 10, no. 1, pp. 54–64, Jan. 2009.

[63] C. E. Rasmussen and C. K. I. Williams, *Gaussian processes for machine learning*, 3. print. Cambridge, Mass.: MIT Press, 2008.

[64] V. Rannou, F. Brouaye, M. Hélier, and W. Tabbara, “Kriging the quantile: application to a simple transmission line model,” *Inverse Probl.*, vol. 18, no. 1, pp. 37–48, Feb. 2002.

[65] F. Jouvie, D. Lecointe, P. Briend, F. Jacquin, and E. Nicolas, “Computation of the field radiated by a FM transmitter by means of ordinary kriging,” *Ann. Telecommun. - Ann. Télécommunications*, vol. 66, no. 7–8, pp. 429–443, Aug. 2011.

[66] O. Jawad, D. Lautru, A. Benlarbi-Delai, J. M. Dricot, and P. De Doncker, “STUDY OF HUMAN EXPOSURE USING KRIGING METHOD,” *Prog. Electromagn. Res. B*, vol. 61, pp. 241–252, 2014.

[67] T. Bdour, C. Guiffaut, and A. Reineix, “Use of Adaptive Kriging Metamodeling in Reliability Analysis of Radiated Susceptibility in Coaxial Shielded Cables,” *IEEE Trans. Electromagn. Compat.*, vol. 58, no. 1, pp. 95–102, Feb. 2016.

[68] H. Bagci, A. C. Yucel, J. S. Hesthaven, and E. Michielssen, “A Fast Stroud-Based Collocation Method for Statistically Characterizing EMI/EMC Phenomena on Complex Platforms,” *IEEE Trans. Electromagn. Compat.*, vol. 51, no. 2, pp. 301–311, May 2009.

[69] O. Aiouaz *et al.*, “Uncertainty analysis of the specific absorption rate induced in a phantom using a stochastic spectral collocation method,” *Ann. Telecommun. - Ann. Télécommunications*, vol. 66, no. 7–8, pp. 409–418, Aug. 2011.

[70] I. S. Stievano, P. Manfredi, and F. G. Canavero, “Parameters Variability Effects on Multiconductor Interconnects via Hermite Polynomial Chaos,” *IEEE Trans. Compon. Packag. Manuf. Technol.*, vol. 1, no. 8, pp. 1234–1239, Aug. 2011.

[71] M. Larbi, I. S. Stievano, F. G. Canavero, and P. Besnier, “Variability Impact of Many Design Parameters: The Case of a Realistic Electronic Link,” *IEEE Trans. Electromagn. Compat.*, vol. 60, no. 1, pp. 34–41, Feb. 2018.

[72] E. Garcia, “Electromagnetic Compatibility Uncertainty, Risk, and Margin Management,” *IEEE Trans. Electromagn. Compat.*, vol. 52, no. 1, pp. 3–10, Feb. 2010.

[73] P. Kersaudy, S. Mostarshedi, B. Sudret, O. Picon, and J. Wiart, “Stochastic Analysis of Scattered Field by Building Facades Using Polynomial Chaos,” *IEEE Trans. Antennas Propag.*, vol. 62, no. 12, pp. 6382–6393, Dec. 2014.

[74] S.-K. Au and J. L. Beck, “Estimation of small failure probabilities in high dimensions by subset simulation,” *Probabilistic Eng. Mech.*, vol. 16, pp. 263–277, Oct. 2001.

[75] T. Konefal, A. C. Marvin, J. F. Dawson, and M. P. Robinson, “A Statistical Model to Estimate an Upper Bound on the Probability of Failure of a System Installed on an Irradiated Vehicle,” *IEEE Trans. Electromagn. Compat.*, vol. 49, no. 4, pp. 840–848, Nov. 2007.

[76] M. Larbi, P. Besnier, and B. Pecqueux, “Probability of EMC Failure and Sensitivity Analysis With Regard to Uncertain Variables by Reliability Methods,” *IEEE Trans. Electromagn. Compat.*, vol. 57, no. 2, pp. 274–282, Apr. 2015.

[77] A. Kouassi, J.-M. Bourinet, S. Lallechere, P. Bonnet, and M. Fogli, “Reliability and Sensitivity Analysis of Transmission Lines in a Probabilistic EMC Context,” *IEEE Trans. Electromagn. Compat.*, vol. 58, no. 2, pp. 561–572, Apr. 2016.

[78] C. Cannamela, J. Garnier, and B. Iooss, “Controlled stratification for quantile estimation,” *Ann. Appl. Stat.*, vol. 2, no. 4, pp. 1554–1580, Dec. 2008.

[79] M. Larbi, P. Besnier, and B. Pecqueux, “The Adaptive Controlled Stratification Method Applied to the Determination of Extreme Interference Levels in EMC Modeling With Uncertain Input Variables,” *IEEE Trans. Electromagn. Compat.*, vol. 58, no. 2, pp. 543–552, Apr. 2016.

[80] J. Shao and D. Tu, *The jackknife and bootstrap*, Second corrected printing. New York: Springer, 1996.

[81] T. Crestaux, O. Le Maıˆtre, and J.-M. Martinez, “Polynomial chaos expansion for sensitivity analysis,” *Reliab. Eng. Syst. Saf.*, vol. 94, no. 7, pp. 1161–1172, Jul. 2009.

[82] A. J. Smola and B. Schölkopf, “A tutorial on support vector regression,” *Stat. Comput.*, vol. 14, no. 3, pp. 199–222, Aug. 2004.

[83] M. Nielsen, *Neural Network and Deep Learning*. 2018.

[84] R. Schobi, B. Sudret, and J. Wiart, “POLYNOMIAL-CHAOS-BASED KRIGING,” *Int. J. Uncertain. Quantif.*, vol. 5, no. 2, pp. 171–193, 2015.

[85] X. Huang, J. Chen, and H. Zhu, “Assessing small failure probabilities by AK–SS: An active learning method combining Kriging and Subset Simulation,” *Struct. Saf.*, vol. 59, pp. 86–95, Mar. 2016.

[86] B. Echard, N. Gayton, M. Lemaire, and N. Relun, “A combined Importance Sampling and Kriging reliability method for small failure probabilities with time-demanding numerical models,” *Reliab. Eng. Syst. Saf.*, vol. 111, pp. 232–240, Mar. 2013.

[87] G. C. Enss, M. Kohler, A. Krzyzak, and R. Platz, “Nonparametric Quantile Estimation Based on Surrogate Models,” *IEEE Trans. Inf. Theory*, vol. 62, no. 10, pp. 5727–5739, Oct. 2016.

[88] R. Trinchero, P. Manfredi, I. S. Stievano, and F. G. Canavero, “Machine Learning for the Performance Assessment of High-Speed Links,” *IEEE Trans. Electromagn. Compat.*, vol. 60, no. 6, pp. 1627–1634, Dec. 2018.

[89] A. Haeri and M. J. Fadaee, “Efficient reliability analysis of laminated composites using advanced Kriging surrogate model,” *Compos. Struct.*, vol. 149, pp. 26–32, Aug. 2016.

[90] G. Vicario, G. Craparotta, and G. Pistone, “Meta-models in Computer Experiments: Kriging versus Artificial Neural Networks: G. VICARIO, G. CRAPAROTTA AND G. PISTONE,” *Qual. Reliab. Eng. Int.*, vol. 32, no. 6, pp. 2055–2065, Oct. 2016.

[91] M. Eldred and J. Burkardt, “Comparison of Non-Intrusive Polynomial Chaos and Stochastic Collocation Methods for Uncertainty Quantification,” 2009.

[92] Y.-J. Kim, “Comparative study of surrogate models for uncertainty quantification of building energy model: Gaussian Process Emulator vs. Polynomial Chaos Expansion,” *Energy Build.*, vol. 133, pp. 46–58, Dec. 2016.

[93] V. Cherkassky and Y. Ma, “Practical selection of SVM parameters and noise estimation for SVM regression,” *Neural Netw.*, vol. 17, no. 1, pp. 113–126, Jan. 2004.

[94] J. A. K. Suykens, J. De Brabanter, L. Lukas, and J. Vandewalle, “Weighted least squares support vector machines: robustness and sparse approximation,” *Neurocomputing*, vol. 48, no. 1–4, pp. 85–105, Oct. 2002.

[95] S. S. Keerthi and C.-J. Lin, “Asymptotic Behaviors of Support Vector Machines with Gaussian Kernel,” *Neural Comput.*, vol. 15, no. 7, pp. 1667–1689, Jul. 2003.

[96] G. F. Smits and E. M. Jordaan, “Improved SVM regression using mixtures of kernels,” 2002, pp. 2785–2790.

[97] J.-M. Bourinet, F. Dehheger, and M. Lemaire, “Assessing small failure probabilities by combined subset simulation and Support Vector Machines,” *Struct. Saf.*, vol. 33, pp. 343–353, Sep. 2011.

[98] T. Hacib, Y. Le Bihan, M. R. Mekideche, H. Acikgoz, O. Meyer, and L. Pichon, “Microwave Characterization Using Least-Square Support Vector Machines,” *IEEE Trans. Magn.*, vol. 46, no. 8, pp. 2811–2814, Aug. 2010.

[99] R. Schöbi, S. Marelli, and B. Sudret, “UQLab user manual – Support vector machines for regression.” 2017.

[100] D. F. Specht, “Probabilistic neural networks,” *Neural Netw.*, vol. 3, no. 1, pp. 109–118, Jan. 1990.

[101] V. Richard, R. Loison, R. Gillard, H. Legay, and M. Romier, “Loss analysis of a reflectarray cell using ANNs with accurate magnitude prediction,” presented at the 11th European Conference on Antennas and Propagation (EUCAP), Paris, France, 2017, pp. 2396–2399.

[102] V. Richard, R. Loison, R. Gillard, H. Legay, and M. Romier, “Optimized Artificial Neural Network for reflectarray cell modelling,” presented at the IEEE International Symposium on Antennas and Propagation (APSURSI), Fajardo, Puerto Rico, 2016, pp. 1211–1212.

[103] A. Djellid, L. Pichon, S. Koulouridis, and F. Bouttout, “MINIATURIZATION OF A PIFA ANTENNA FOR BIOMEDICAL APPLICATIONS USING ARTIFICIAL NEURAL NETWORKS,” *Prog. Electromagn. Res. M*, vol. 70, pp. 1–10, 2018.

[104] F. Burghardt, S. Sandmann, and H. Garbe, “Simplifying Risk Analysis to Determine the Influence of Wind Turbines to the Electric Field of a DVOR Antenna Using Artificial Neural Networks,” in *2018 International Symposium on Electromagnetic Compatibility (EMC EUROPE)*, Amsterdam, 2018, pp. 283–287.

[105] H. Kim, C. Sui, K. Cai, B. Sen, and J. Fan, “An Efficient High-Speed Channel Modeling Method Based on Optimized Design-of-Experiment (DoE) for Artificial Neural Network Training,” *IEEE Trans. Electromagn. Compat.*, pp. 1–7, 2018.

[106] D. Shi and Y. Gao, “A New Method for Identifying Electromagnetic Radiation Sources Using Backpropagation Neural Network,” *IEEE Trans. Electromagn. Compat.*, vol. 55, no. 5, pp. 842–848, Oct. 2013.

[107] H. Haralambous and H. Papadopoulos, “24-h HF Spectral Occupancy Characteristics and Neural Network Modeling Over Northern Europe,” *IEEE Trans. Electromagn. Compat.*, vol. 59, no. 6, pp. 1817–1825, Dec. 2017.

[108] M. K. Smail, Y. Le Bihan, and L. Pichon, “Fast diagnosis of transmission lines using neural networks and principal component analysis,” *Int. J. Appl. Electromagn. Mech.*, vol. 39, no. 1–4, pp. 435–441, Sep. 2012.

[109] M. T. Hagan and M. B. Menhaj, “Training feedforward networks with the Marquardt algorithm,” *IEEE Trans. Neural Netw.*, vol. 5, no. 6, pp. 989–993, Nov. 1994.

[110] F. Dan Foresee and M. T. Hagan, “Gauss-Newton approximation to Bayesian learning,” in *Proceedings of International Conference on Neural Networks (ICNN’97)*, Houston, TX, USA, 1997, vol. 3, pp. 1930–1935.

[111] N. Wiener, “The Homogeneous Chaos,” *Am. J. Math.*, vol. 60, no. 4, p. 897, Oct. 1938.

[112] D. Xiu, *Numerical methods for stochastic computations: a spectral method approach*. Princeton, NJ: Princeton Univ. Press, 2010.

[113] A. O’Hagan, “Polynomial Chaos: A Tutorial and Critique from a Statisticianís Perspective.”

[114] G. Blatman and B. Sudret, “Adaptive sparse polynomial chaos expansion based on least angle regression,” *J. Comput. Phys.*, vol. 230, no. 6, pp. 2345–2367, Mar. 2011.

[115] Q. Shao, A. Younes, M. Fahs, and T. A. Mara, “Bayesian sparse polynomial chaos expansion for global sensitivity analysis,” *Comput. Methods Appl. Mech. Eng.*, vol. 318, pp. 474–496, May 2017.

[116] P. Kersaudy, S. Mostarshedi, B. Sudret, O. Picon, and J. Wiart, “Stochastic Analysis of Scattered Field by Building Facades Using Polynomial Chaos,” *IEEE Trans. Antennas Propag.*, vol. 62, no. 12, pp. 6382–6393, Dec. 2014.

[117] T. Bdour and A. Reineix, “Global Sensitivity Analysis and Uncertainty Quantification of Radiated Susceptibility in PCB Using Nonintrusive Polynomial Chaos Expansions,” *IEEE Trans. Electromagn. Compat.*, vol. 58, no. 3, pp. 939–942, Jun. 2016.

[118] S. Marelli and B. Sudret, “UQLab user manual – Polynomial Chaos Expansions.” 2017.

[119] R. Schöbi, S. Marelli, and B. Sudret, “UQLab user manual – Kriging (Gaussian process modelling).” 2017.

[120] W. Ye and F. You, “A computationally efficient simulation-based optimization method with region-wise surrogate modeling for stochastic inventory management of supply chains with general network structures,” *Comput. Chem. Eng.*, vol. 87, pp. 164–179, Apr. 2016.

[121] Y. Zhao, W. Lu, and C. Xiao, “A Kriging surrogate model coupled in simulation optimization approach for identifying release history of groundwater sources,” *J. Contam. Hydrol.*, vol. 185–186, pp. 51–60, Feb. 2016.

[122] T. Mukhopadhyay, S. Naskar, S. Dey, and S. Adhikari, “On quantifying the effect of noise in surrogate based stochastic free vibration analysis of laminated composite shallow shells,” *Compos. Struct.*, vol. 140, pp. 798–805, Apr. 2016.

[123] W. Valente, A. Dalla’Rosa, A. Raizer, and L. Pichon, “The Use of TLM and Kriging Methods for Electromagnetic Compatibility Management in Health Care Facilities,” *IEEE Trans. Magn.*, vol. 44, no. 6, pp. 1478–1481, Jun. 2008.

[124] S. Koziel, S. Ogurtsov, I. Couckuyt, and T. Dhaene, “Efficient simulation-driven design optimization of antennas using co-kriging,” 2012, pp. 1–2.

[125] B. Xia, S. Hong, K. Choi, and C. S. Koh, “Optimal Design of Winding Transposition of Power Transformer Using Adaptive Co-Kriging Surrogate Model,” *IEEE Trans. Magn.*, vol. 53, no. 6, pp. 1–4, Jun. 2017.

[126] T. Bdour, C. Guiffaut, and A. Reineix, “EFFICIENT BAYESIAN MULTIFIDELITY APPROACH IN METAMODELING OF AN ELECTROMAGNETIC SIMULATOR RESPONSE,” *Prog. Electromagn. Res. M*, vol. 54, pp. 47–54, 2017.

[127] P. Goovaerts, “Geostatistical Interpolation of Rate Data Using Poisson Kriging,” in *Encyclopedia of GIS*, S. Shekhar, H. Xiong, and X. Zhou, Eds. Cham: Springer International Publishing, 2015, pp. 1–8.

[128] J. Pilz and G. Spöck, “Why do we need and how should we implement Bayesian kriging methods,” *Stoch. Environ. Res. Risk Assess.*, vol. 22, no. 5, pp. 621–632, Aug. 2008.

[129] V. Dubourg, B. Sudret, and J.-M. Bourinet, “Reliability-based design optimization using kriging surrogates and subset simulation,” *Struct. Multidiscip. Optim.*, vol. 44, no. 5, pp. 673–690, Nov. 2011.

[130] P. Kersaudy, B. Sudret, N. Varsier, O. Picon, and J. Wiart, “A new surrogate modeling technique combining Kriging and polynomial chaos expansions – Application to uncertainty analysis in computational dosimetry,” *J. Comput. Phys.*, vol. 286, pp. 103–117, Apr. 2015.

[131] R. Schöbi, B. Sudret, and S. Marelli, “Rare Event Estimation Using Polynomial-Chaos Kriging,” *ASCE-ASME J. Risk Uncertain. Eng. Syst. Part Civ. Eng.*, vol. 3, no. 2, p. D4016002, Jun. 2017.

[132] R. Schöbi, S. Marelli, and B. Sudret, “UQLab user manual – PC-Kriging.” 2017.

[133] T. Houret, P. Besnier, S. Vauchamp, and P. Pouliguen, “Combining Kriging and Controlled Stratification

to Identify Extreme Levels of Electromagnetic Interference," *EMC Europe*, Barcelona, 2019.

[134] G. Blatman and B. Sudret, "An adaptive algorithm to build up sparse polynomial chaos expansions for stochastic finite element analysis," *Probabilistic Eng. Mech.*, vol. 25, no. 2, pp. 183–197, Apr. 2010.

[135] J. Zhang and A. A. Taflanidis, "Adaptive Kriging Stochastic Sampling and Density Approximation and Its Application to Rare-Event Estimation," *ASCE-ASME J. Risk Uncertain. Eng. Syst. Part Civ. Eng.*, vol. 4, no. 3, p. 04018021, Sep. 2018.

[136] M. A. Bouhlel, N. Bartoli, A. Otsmane, and J. Morlier, "Improving kriging surrogates of high-dimensional design models by Partial Least Squares dimension reduction," *Struct. Multidiscip. Optim.*, vol. 53, no. 5, pp. 935–952, May 2016.

[137] B. W. Silverman, *Density estimation for statistics and data analysis*, 1., CRC Press repr. Boca Raton, Fla.: Chapman & Hall/CRC, 1986.

[138] A. Olsson, G. Sandberg, and O. Dahlblom, "On Latin hypercube sampling for structural reliability analysis," *Struct. Saf.*, vol. 25, no. 1, pp. 47–68, Jan. 2003.

[139] J. M. Hilbe, *Negative binomial regression*, 2. ed., repr. With corr. Cambridge [u.al.]: Cambridge Univ. Press, 2011.

[140] I. M. Sobol′, "Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates," *Math. Comput. Simul.*, vol. 55, no. 1–3, pp. 271–280, Feb. 2001.

[141] S. Greenland *et al.*, "Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations," *Eur. J. Epidemiol.*, vol. 31, no. 4, pp. 337–350, Apr. 2016.

[142] R. Sanchez-Iborra, M.-D. Cano, and J. Garcia-Haro, "Performance evaluation of QoE in VoIP traffic under fading channels," in *2013 World Congress on Computer and Information Technology (WCCIT)*, Sousse, Tunisia, 2013, pp. 1–6.

[143] V. M. Artyushenko and V. I. Volovach, "Nakagami Distribution Parameters Comparatively Estimated by the Moment and Maximum Likelihood Methods," *Optoelectron. Instrum. Data Process.*, vol. 55, no. 3, pp. 237–242, May 2019.

[144] K. Ahmad, S. P. Ahmad, and A. Ahmed, "Classical and Bayesian Approach in Estimation of Scale Parameter of Nakagami Distribution," *J. Probab. Stat.*, vol. 2016, pp. 1–8, 2016.

[145] S. A. Ahmed, "Comparative study of four methods for estimating Weibull parameters for Halabja, Iraq," *Int. J. Phys. Sci.*, vol. 8, no. 5, pp. 186–192, 2013.

[146] T. Houret, "Catalogue of Usefull Matlab Functions," 2019.

[147] M. Leone, "Closed-Form Expressions for the Electromagnetic Radiation of Microstrip Signal Traces," *IEEE Trans. Electromagn. Compat.*, vol. 49, no. 2, pp. 322–328, May 2007.

ANNEXES

I. Annexes du Chapitre II

1. Inférence par méthode des moments

La méthode des moments se base sur les moments statistiques d'ordre r notés μ_r, d'une v.a. S :

$$\mu_r = E[S^r]. \tag{65}$$

Le moment d'ordre 1 est l'espérance, tandis que la variance se déduit de celui d'ordre 2 :

$$V[S] = E[S^2] - E[S]^2 = \mu_2 - \mu_1^2. \tag{66}$$

Les moments μ_r sont liés aux paramètres $\boldsymbol{\theta}$ par la fonction g_r, fonction génératrice des moments. Les moments, généralement inconnus peuvent être estimés par $\widehat{\mu_r}$, à partir de l'échantillon. Il suffit d'estimer autant de moments que de composantes de $\boldsymbol{\theta}$. On obtient ainsi un système comportant autant d'équations que d'inconnues :

$$\widehat{\mu_r} = \frac{1}{n}\sum_{e=1}^{n} S_e^r = g_r(\boldsymbol{\theta}) \;\; r = \{1, \dots, Card(\boldsymbol{\theta})\}. \tag{67}$$

Table 1 : Fonction génératrice des moments selon la loi sous-jacente.

Loi	N	LN	W
$\boldsymbol{\theta_1}$	μ	a	γ
$\boldsymbol{\theta_2}$	σ	b	β
$g_r(\boldsymbol{\theta})$	$e^{r\mu+\frac{r^2\sigma}{2}}$	$e^{ra+\frac{r^2 b}{2}}$	$\gamma^r \Gamma\left(1+\frac{r}{\beta}\right)$

Résoudre le système d'équation (67) permet d'obtenir l'estimateur le plus probable de $\boldsymbol{\theta}$. Ceux-ci sont données dans Table 2 pour les lois N et LN [38] ainsi que W [145]. Pour le cas de W, ce sont des approximations mais qui peuvent être utilisées comme point de départ pour un algorithme de recherche itérative (fonction Matlab *vpasolve*).

Table 2 : Estimateur le plus probable des paramètres de $\boldsymbol{\theta}$ d'après la méthode des moments selon la loi sous-jacente.

Loi	N(μ,σ)	LN(a,b)	W(γ, β)
$\hat{\theta}_1$	$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} S_i$	$\hat{a} = ln(\hat{\mu}) - \frac{1}{2}\hat{b}^2$	$\hat{\gamma} \approx \frac{\hat{\mu}}{\Gamma\left(1+\frac{1}{\hat{\beta}}\right)}$
$\hat{\theta}_2$	$\widehat{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n}(S_i - \hat{\mu})^2$	$\hat{b} = \sqrt{ln\left(1+\left(\frac{\hat{\sigma}}{\hat{\mu}}\right)^2\right)}$	$\hat{\beta} \approx \left(\frac{\hat{\sigma}}{\hat{\mu}}\right)^{-1.086}$

2. Implémentation de l'inférence bayésienne en Matlab

Un catalogue rassemblant des fonctions Matlab utiles se trouve dans [146]. De plus, nous fournissons ici la fonction Matlab implémentant l'inférence bayésienne de l'équation (5) page 23 :

```
function [ fpost,Ls,Us ] = IB(data,s,theta,loi,priori,res)
%{
Fonction implémentant l'inférence Bayésienne

Entrées de la fonction :
data: seuils de susceptibilités (vecteur colonne)
theta: paramètres à tester (matrice 1 colonne par composante)
loi: nom de la distribution (string)
priori: densité a priori de theta (fonction du type [ftheta]=priori(theta))
s: seuils de susceptibilité possibles (vecteur)

Sorties de la fonction :
fpost: densité a posteriori de theta (matrice)
Ls: lower bound de la probabilité de défaillance (vecteur)
Us: upper bound de la probabilité de défaillance (vecteur)
%}

% Definition de la pdf f_Like
If size(data,2)==1
        like=@(data,theta) prod(...
                                pdf(loi,...
                                        repmat(data.',length(theta),1),...
                                        repmat(theta(:,1),1,length(data)),...
                                        repmat(theta(:,2),1,length(data))),...
                                2);
elseif size(data,2)==1
        like=@(data,theta) ...
                prod(...
                        cdf(loi,...
                                repmat(data(:,2).',length(theta),1),...
                                repmat(theta(:,1),1,length(data)),...
                                repmat(theta(:,2),1,length(data)))...
                        -cdf(loi,...
                                repmat(data(:,1).',length(theta),1),...
                                repmat(theta(:,1),1,length(data)),...
                                repmat(theta(:,2),1,length(data))),...
                2);
else
    error('mauvais formatage de Data')
end
pf=@(s,theta) cdf(loi,...
                repmat(s,length(theta),1),...
                repmat(theta(:,1),1,length(s)),...
                repmat(theta(:,2),1,length(s)));
% Calcul de la densité a posteriori
evid=priori(theta).*like(data,theta);
fpost=evid./sum(evid);
% Calcul de la distribution de la probabilité de défaillance et de ses bornes
pfs=pf(s,theta);
[pfs_sort,ind]=sort(pfs);
```

```
fpost_cum=cumsum(fpost(ind));
[~,ind_sup]=max(fpost_cum>=0.975);
[~,ind_inf]=max(fpost_cum>=0.275);
Ls=diag(pfs_sort(ind_inf,:)).';
Us=diag(pfs_sort(ind_sup,:)).';
end
```

II. Annexes du Chapitre III

1. Calcul du modèle PCB

Les calculs sont issus en majeur partie de [50]. La sortie se calcule ainsi :

$$Y = \frac{\omega\mu_0}{4\pi}\frac{e^{-jk_0r}}{r}\frac{V_S}{Z_S+Z_C}\frac{1}{1-\rho_S\rho_L e^{-j\beta 2l}}\left[\boldsymbol{F}_x\left(\frac{1-e^{-j(\beta-k_x)l}}{\beta-k_x}+\rho_L e^{-j\beta 2l}\frac{1-e^{j(\beta+k_x)l}}{\beta+k_x}\right)+jh\boldsymbol{F}_z(1-\rho_l e^{-2j\beta}-(1-\rho_L)e^{-j(\beta-k_x)l}\right], \tag{68}$$

où :

$$\boldsymbol{F}_x \approx 2jk_0h\left[\left(\frac{sin^2\theta}{\epsilon_r}-1\right)cos\phi e_\theta cos\theta sin\phi e_\phi\right], \tag{69}$$

$$\boldsymbol{F}_z \approx \frac{2\,sin\,\theta e_\theta}{\epsilon_r}, \tag{70}$$

$$k_0 = \frac{\omega}{c_0}, \tag{71}$$

$$k_x = k_0 sin\theta cos\phi, \tag{72}$$

$$\beta = k_0\sqrt{\epsilon_{eff}}, \tag{73}$$

$$\rho_{S,L} = \frac{Z_{S,L}-Z_C}{Z_{S,L}+Z_C}. \tag{74}$$

De plus, l'impédance caractéristique Z_C se calcule ainsi :

$$Z_C = \frac{120\pi}{2\pi\sqrt{2}\sqrt{\epsilon_{eff}+1}}ln\left(1+\frac{4h}{w_p}\left(a+\sqrt{a^2+d}\right)\right), \tag{75}$$

où :

$$a = \frac{14+\frac{8}{\epsilon_{eff}}}{11}\frac{4h}{W_p}, \tag{76}$$

$$d = 0.5\left(1+\frac{1}{\epsilon_{eff}}\right)\pi^2 \tag{77}$$

$$W_p = W + \Delta_{W_p}, \tag{78}$$

$$\Delta_{W_p} = \Delta_W\left(1+\frac{1}{\epsilon_{eff}}\right)\frac{1}{2}, \tag{79}$$

La permittivité diélectrique effective ϵ_{eff} se calcule ainsi :

$$\epsilon_{eff} = \frac{1}{2}(\epsilon_r+1)+\frac{1}{2}(\epsilon_r-1)\left(1+\frac{10h}{W}\right)^{-a_1.b_1}, \tag{80}$$

$$a_1 = 1-\frac{1}{49}ln\left(\frac{u^4+\left(\frac{u}{52}\right)^2}{u^4+0.432}\right)+ln\left(1+\left(\frac{u}{18.1}\right)^3\right)\frac{1}{18.7}, \tag{81}$$

$$u = \frac{W}{h}, \tag{82}$$

$$b_1 = 0.564\frac{(\epsilon_r-0.9)^{0.053}}{\epsilon_r+3}. \tag{83}$$

2. Implémentation de la stratification contrôlée en Matlab

```
function [ qSC ,F_SC,ySC,Output_strates,qM ,n_strates] = SC(nSC,yM,strates_proba,upperTail,mySM,Input)
%{
Fonction implémentant la Stratification Contrôlée (SC), le métamodèle étant déjà construit par UQLab. Cette
fonction nécessite les fonctions « DOE » et « extremSM ».

Entrées de la fonction :
nSC : budget pour la SC (scalaire).
yM : sorties de grande taille, e.g. 1 million, prédites par le métamodèle (vecteur).
strates_proba : probabilités définissant les limites des strates (vecteur).
upperTail : booléen égal à 1 si  quantile de probabilité proche 1, 0 si la probabilité est proche de 0 (scalaire).
mySM : métamodèle support de la SC (UQLab object).
Input : variables aléatoires du modèle de contrainte (UQLab object).

Sorties de la fonction :
qSC : quantile estimée par la SC (scalaire).
F_SC : cdf estimée (vecteur).
ySC : points où F_SC a été calculé (vecteur).
Output_strates : réalisations de sortie, rangées par strates (cell array).
qM : quantiles estimés par le métamodèle (vecteur).
n_strates : répartition du budget accordée SC par strates (vecteur).
%}

%% section répartissant  le budget par strate (allocation uniforme ici)
if upperTail==0
        strates_proba=[0,proba_cible,2*proba_cible,.5,1];
        n_strates=floor(nSC*ones(1,length(strates_proba)-1)./(length(strates_proba)-1));
        n_strates(1)=n_strates(1)+mod(nSC,length(n_strates));
elseif upperTail==1
        strates_proba=[0,0.5,1-2*(1-proba_cible),proba_cible,1];
        n_strates=floor(nSC*ones(1,length(strates_proba)-1)./(length(strates_proba)-1));
        n_strates(end)=n_strates(end)+mod(nSC,length(n_strates));
end
%%
qM=quantile(yM,strates_proba) ;
if length(unique(qM))==length(strates_proba) && sum(isnan(qM))==0

%% section calculant les potentielles valeurs extrêmes
        Output_strates=cell(1,length(n_strates));
        for j=1:length(n_strates)
                x_new= extremSM( mySM,n_strates(j),strates_proba(j:j+1),Input);
                disp('Début de la simulation HFSS de la strate numero:')
                j
                Output_strates(j)= {DOE(x_new)};
        end
%% section calculant la cdf de l'échantillon stratifié, et donc le quantile ciblé
        yind_unif=Output_strates;
        yind_v_unif=cell2mat((yind_unif).');
        FySC_unif=zeros(size(yind_v_unif));
        pj_unif=FySC_unif;
        alpha=diff(strates_proba);
        for i=1:length(n_strates)
                temp=cell2mat(yind_unif(i));
                for j=1:length(yind_v_unif)
                        pj_unif(j)=sum(temp<=yind_v_unif(j));
```

```
                end
                pj_unif=pj_unif*alpha(i)/n_strates(i);
                FySC_unif=FySC_unif+pj_unif;
        end
        [F_SC,inds_unif]=sort(FySC_unif);
        ySC=yind_v_unif(inds_unif);
        ind_quantile=find(F_SC>proba_cible,1,'first');
        qSC=ySC(ind_quantile);
 %%

elseif length(unique(qM))<length(strates_proba)
        if upperTail==0
                qSC=qM(1);
        else
                qSC=qM(end);
        end
        warning('Taille initial du DOE trop petite: SM très imprécis)

elseif sum(isnan(qM))==length(strates_proba)
        qSC=NaN(1);
        warning('Taille initial du DOE trop petite: SM pas construit correctement!)
end
end

function [ x_xtm,y_xtm ] = extremSM( mySM,n_ext,alpha,Input)
%{
Fonction permettant de trouver les valeurs d'entrée générant des valeurs extrêmes de sortie selon le
métamodèle. Ces points candidats seront ensuite simulés par HFSS par la fonction « DOE ».

Entrées de la fonction :
mySM : métamodèle support de la SC (UQLab object).
n_ext : nombre de candidats à trouver (scalaire).
alpha : probabilité du quantile ciblé (scalaire).
Input : variables aléatoires du modèle de contrainte (UQLab object).

Sorties de la fonction :
 x_xtm : valeurs des variables d'entrée candidats (vecteur).
 y_xtm : valeurs de sortie prédite par le métamodèle (vecteur).
%}

p=1-max(alpha(alpha<1));
SampleSize=n_ext+icdf('Negative Binomial',.95,1,p);
x=zeros(SampleSize,length(Input.Marginals));
x_xtm=[];
y_xtm=[];
while size(x_xtm,1)<=n_ext
        size(x_xtm,1);
        for i=1:length(Input.Marginals)
                x(:,i)=random(Input.Marginals(i).dist,SampleSize,1);
        end
        y = uq_evalModel(mySM, x);
        qSM=quantile(y,alpha);
        ind_xtm=find(y>qSM(1) & y<=qSM(2));
        x_xtm=[x_xtm;x(ind_xtm,:)];
        y_xtm=[y_xtm;y(ind_xtm)];
end
r=randperm(size(x_xtm,1),n_ext);
```

```
x_xtm=x_xtm(r,:);
y_xtm=y_xtm(r);
end
```

La fonction « *DOE* » construit le DOE en faisant appel à HFSS. L'implémentation de la fonction « *DOE* » fournie ci-dessous doit être légèrement adaptée au cas étudié. Cette fonction écrit les valeurs de *X_DOE* dans un fichier csv (ici « param_sweep.csv »). Puis, elle lance l'exécution d'un script python (ici « cas_test.py ») qui permet la simulation HFSS du modèle de contrainte avec les paramètres spécifiées par *X_DOE*. Ce script écrit les résultats dans un fichier csv (ici « param_sweep_result.csv ») qui est ensuite lu par Matlab.

```
function [ Y ] = DOE(X_DOE)
%{
Fonction construisant le DOE grâce à HFSS. A adapter au cas d'étude :
-Noms et unités des variables  qui doivent correspondre exactement à ceux utilisés par HFSS
-Nom des fichiers « param_sweep.csv » et « cas_test.py » ainsi que leur chemin « PATH ».

Entrées de la fonction :
X_DOE : réalisations d'entrée du DOE (matrice où les colonnes= variables, et les lignes = réalisations).

Sorties de la fonction :
Y : réalisations de sortie du DOE (vecteur).
%}

X_names={'hauteur','longueur','rayon','rayonC1','rotX','rotY',...
   'rotZ','xGND','yGND'};
X_units=cell(1,length(X_names));
X_units([1:4,8:9])={'(mm)'};
X_units(5:7)={'(deg)'};
header=strjoin(strcat(X_names,X_units),',');
dlmwrite('param_sweep.csv',header,'Delimiter','');
dlmwrite('param_sweep.csv',X_DOE,'-append','Delimiter',',');

system('"PATH\...\ansysedt.exe" -RunScriptAndExit "cas_test.py"');

fileIDR = fopen('param_sweep_results.csv');
colonnes=cell(1,3+length(X_names)+length(X_units));
colonnes(2:2:end)={'%n'};
colonnes(3:2:end)={'%s'};
colonnes(1)={'%n'};
C = textscan(fileIDR,strjoin(colonnes),'Delimiter',',','HeaderLines',1);
fclose(fileIDR);
Y=C{end-1};
end
```

Le script python « cas_test.py » peut être généré automatiquement par HFSS (Tools/Record script to File). Chaque action par l'interface graphique est ainsi traduite automatiquement en ligne de commande. De même que la fonction « DOE », ce script doit être adapté au cas étudié. A titre d'exemple nous fournissons le script python correspondant au chapitre IV où le fichier HFSS s'appelle « cas_testCetS.aedt » situé dans « PATH ».

```
# ----------------------------------------------
# Script Recorded by ANSYS Electronics Desktop Version 2015.2.0
# 18:12:03  sept. 26, 2019
# ----------------------------------------------
import ScriptEnv
ScriptEnv.Initialize("Ansoft.ElectronicsDesktop")
```

```
oDesktop.RestoreWindow()
oDesktop.OpenProject("PATH/cas_test_CetS.aedt")
oProject = oDesktop.SetActiveProject("cas_test_CetS")
oDesign = oProject.SetActiveDesign("HFSSDesign1")
oModule = oDesign.GetModule("Optimetrics")
oModule.ImportSetup("OptiParametric",
	[
		"NAME:ParametricSetup1",
		"PATH\\param_sweep.csv"
	])
oModule.EditSetup("ParametricSetup1",
	[
		"NAME:ParametricSetup1",
		"IsEnabled:="		, True,
		[
			"NAME:ProdOptiSetupData",
			"SaveFields:="		, False,
			"CopyMesh:="		, False
		],
		[
			"NAME:StartingPoint"
		],
		"Sim. Setups:="		, ["Setup1"]
		,
		[
			"NAME:Goals",
			[
				"NAME:Goal",
				"ReportType:="		, "Fields",
				"Solution:="		, "Setup1 : LastAdaptive",
				[
					"NAME:SimValueContext"
				],
				"Calculation:="		, "mag(I)",
				"Name:="		, "mag(I)",
				[
					"NAME:Ranges",
					"Range:="		, [
		"Var:="		, "Freq",		"Type:="
	, "d",		"DiscreteValues:="		, "0.868GHz"],
					"Range:="		, [
		"Var:="		, "Phase",
	"Type:="		, "d",		"DiscreteValues:="		,
"0deg"]
				]
			]
		]
	])
oProject.Save()
oDesign.AnalyzeAll()
oModule.ExportOptimetricsResult("ParametricSetup1", "PATH/param_sweep_results.csv", False)
oProject.Save()
oModule.DeleteSetups(["ParametricSetup1"])
oDesign.DeleteFullVariation("All", False)
oProject.Save()
oDesktop.CloseProject("cas_test_CetS")
```

III. Liste des Figures et des Tableaux

1. Liste des Figures

Figure 1 : Le NEROD F5 de MC2 technologies. Cette arme permet de générer une IEMI ciblée afin de forcer l'atterrissage d'un drone en toute sécurité. Photo datant du défilé militaire du 14 juillet 2019 à Paris. 11

Figure 2 : Schéma résumant la problématique d'une étude de risque en cas d'IEMI : les incertitudes épistémiques (termes associés aux points d'interrogation) impliquent une contrainte (rouge) et une susceptibilité (verte) incertaines. 13

Figure 3 : Schéma résumant notre stratégie pour évaluer la probabilité de défaillance d'équipements soumis à une agression de type IEMI. Les encadrés (bleu, vert et rouge) correspondent aux trois étapes de notre stratégie. Les termes soulignés correspondent aux solutions techniques mises en place..... 15

Figure 4 : Schéma de principe des tests de susceptibilité équipements. Etat haut = défaillance..... 19

Figure 5: Effet des paramètres $\boldsymbol{\theta} = [\gamma, \beta]$ sur la pdf d'une loi de W. 21

Figure 6 : Effet des paramètres $\boldsymbol{\theta} = [\mu, \sigma 2]$ sur la pdf d'une loi N..... 21

Figure 7 : Ensemble non exhaustif des distributions univariées existantes et leurs liens relatifs (soit par leurs paramètres, soit par transformation de variable). Figure extraite de [29]. Encadrés rouges = distributions utilisées dans ce chapitre..... 22

Figure 8 : Distance moyenne (en absolue) entre les trois distributions..... 28

Figure 9 : Probabilité de choisir la loi N si la distribution vraie est effectivement N. 31

Figure 10 : Probabilité de choisir la loi LN (resp. W) alors que la loi vraie est LN (resp. W)..... 31

Figure 11 : Loi Normale, estimation du paramètre $\theta 1$..... 32

Figure 12 : Loi Normale, estimation du paramètre $\theta 2$..... 33

Figure 13 : Loi W, estimation du paramètre $\theta 1$..... 33

Figure 14 : Loi W, estimation du paramètre $\theta 2$..... 34

Figure 15 : Illustration du caractère aléatoire des bornes à l'aide de 5 réalisations différentes de taille n et des estimations de L et U correspondantes..... 35

Figure 16 : Loi N, distributions des bornes s_{low} et s_{upp} avec un intervalle de confiance à 95 % encadrant le quantile 5 % avec une taille d'échantillon de n=100..... 36

Figure 17 : Loi W, distributions des bornes s_{low} et s_{upp} avec un intervalle de confiance à 95 % encadrant le quantile 5 % avec une taille d'échantillon de n=100..... 36

Figure 18 : Loi N distribution des bornes s_{low} et s_{upp} avec un intervalle de confiance à 95 % encadrant le quantile 5 % avec une taille d'échantillon de n=5. 37

Figure 19 : Loi W, distributions des bornes s_{low} et s_{upp} avec un intervalle de confiance à 95 % encadrant le quantile 5 % avec une taille d'échantillon de n=5..... 37

Figure 20 : Loi N, moyenne et écart type des bornes s_{low} et s_{upp} encadrant le quantile 5 % en fonction de la taille d'échantillon. 38

Figure 21 : Loi W, moyenne et écart type des bornes s_{low} et s_{upp} encadrant le quantile 5 % en fonction de la taille d'échantillon. 39

Figure 22 : Equipment électronique virtuel..... 40

Figure 23 : Ensemble des seuils simulés constituant l'échantillon de référence et ajustements au sens IMV des lois N, LN et W..... 41

Figure 24 : Equipement virtuel. Encadrement probabiliste du quantile 5 % de la distribution de susceptibilité en fonction de la taille d'échantillon (s_{low} et s_{upp}). 43

Figure 25 : Equipement virtuel. Échantillons de taille n=5, distributions des bornes (s_{low} et s_{upp}), calculées avec une confiance de 95 % encadrant le quantile 5 %..... 43

Figure 26 : Equipment virtuel. Échantillons de taille n=100, distributions des bornes (s_{low} et s_{upp}), calculées avec

une confiance de 95 % encadrant le quantile 5 %. 44

Figure 27 : Famille d'alimentations A : mesures, U(s) et L(s) ainsi que $pf(s)$ le plus probable. 46

Figure 28 : Procédure d'estimation de la distribution de susceptibilité par la méthode de Bootstrap. 48

Figure 29 : Exemple d'utilisation d'un SVM dans le cadre d'une régression linéaire. Seuls les points xj et xi, qui sont en dehors du tube (défini ici par les deux droites pointillées noires) sont pénalisés. 54

Figure 30 : Exemple d'architecture de NN avec deux variables d'entrée (X_1 et X_2), une de sortie (Y_1) et deux couches intermédiaires. Tous les poids et les biais ne sont pas représentés. 55

Figure 31 : Schéma de principe de la stratification contrôlée dans le cas où n_s=4, N_j=2 $\forall\, j$ et $\boldsymbol{A} = 0, \alpha, 2\alpha, 0.5, 1$. 60

Figure 32 : Schéma de la piste PCB (extrait de [133]). 62

Figure 33 : Modèle de rayonnement PCB : histogramme des 10^6 réponses du modèle vrai du rayonnement des PCB et des quantiles réels qui seront estimés. 63

Figure 34 : Modèle de rayonnement PCB : champ électrique rayonné pour 10000 entrées aléatoires (points rouges pour $Y>y_{90\%}$) et variation du champ électrique rayonné en fonction de la fréquence tandis que les autres entrées sont fixées à leur valeur nominale. 63

Figure 35 : Modèle de circuit RLC série. 64

Figure 36 : Modèle S_{11} : histogramme des 10^6 réponses du modèle vrai et des quantiles vrais qui seront estimés. 65

Figure 37 : Modèle S_{11} : 10000 réalisations de sortie (en linéaire). Les points rouges représentent des sorties considérées comme extrêmes i.e. inférieures à un quantile (ici 10 %). La courbe noire correspond au cas où toutes les entrées sont fixées à leur valeur nominale à l'exception de la fréquence. 65

Figure 38 : Modèle PCB : distributions d'erreurs généralisées pour les valeurs extrêmes ($Y>y_{90\%}$) en fonction des tailles de DOE (n). 67

Figure 39 : Modèle PCB : distribution de la corrélation pour un DOE de taille n=50 en fonction de la probabilité α. 68

Figure 40 : Modèle S11 (RLC réflexion): distributions d'erreurs généralisées pour les valeurs extrêmes ($Y\leq y_{10\%}$) en fonction des tailles de DOE (n). 69

Figure 41 : Modèle RLC : distribution de la corrélation pour un DOE de taille n=50 en fonction de la probabilité α (1 %, 5 % et 9 %). 69

Figure 42 : Modèle S11, estimation de $q_{10\%}$. Erreur relative moyenne (θ_1) par estimation empirique (EE), krigeage (KRI) et krigeage + Stratification Contrôlée (K-SC). Les résultats pour θ_2 se trouvent dans la Figure 43. 72

Figure 43 : Modèle S11, estimation de $q_{10\%}$. Erreur relative absolue maximale (θ_2) par estimation empirique (EE), krigeage (KRI) et krigeage + Stratification Contrôlée (K-SC). Les résultats pour θ_1 se trouvent dans la Figure 42. 72

Figure 44 : Modèle S11, estimation de $q_{1\%}$. Erreur relative moyenne ($\theta 1$) par estimation empirique (EE), krigeage (KRI) et krigeage + Stratification Contrôlée (K-SC). Les résultats pour θ_2 se trouvent dans la Figure 45. 73

Figure 45 : Modèle S11, estimation de $q_{1\%}$. Erreur relative absolue maximale ($\theta 2$) par estimation empirique (EE), krigeage (KRI) et krigeage + Stratification Contrôlée (K-SC). Les résultats pour θ_1 se trouvent dans la Figure 44. 73

Figure 46 : Estimation du rayonnement des PCB $q_{90\%}$: erreur relative (%) par estimation empirique (EE), krigeage (KRI) et krigeage + Stratification Contrôlée (K-SC). Les courbes « smooth » sont obtenue par filtrage des points (moyenne par fenêtre glissante). 75

Figure 47 : Estimation du rayonnement des PCB $q_{99\%}$: erreur relative (%) par estimation empirique (EE), krigeage (KRI) et krigeage + Stratification Contrôlée (K-SC). Les courbes « smooth » sont obtenue par filtrage des points (moyenne par fenêtre glissante). 75

Figure 48 : Estimation du rayonnement des PCB $q_{99,9\%}$: erreur relative (%) par estimation empirique (EE), krigeage (KRI) et krigeage + Stratification Contrôlée (K-SC). Les courbes « smooth » sont obtenue par filtrage

des points (moyenne par fenêtre glissante). 76

Figure 49 : Schéma de l'algorithme d'implémentation pratique de K-SC. 78

Figure 50 : Diagramme de l'algorithme pour valider partie SM de la Figure 49 (aussi en bleu). 82

Figure 51 : Modèle S11 : évolution de la moyenne de l'estimation du quantile à 1 % en fonction de la taille du DOE. Les points représentent les évaluations de l'algorithme de la Figure 49. Chaque couleur correspond à une évaluation de l'algorithme. La courbe représente la pseudo-référence issue de l'algorithme en Figure 50 83

Figure 52 : Modèle S11 : évolution de l'écart type de l'estimation du quantile 1 % en fonction de la taille du DOE. Les points représentent les évaluations de l'algorithme de la Figure 49. Chaque couleur correspond à une évaluation de l'algorithme. La courbe représente la pseudo-référence issue de l'algorithme en Figure 50 84

Figure 53 : Modèle PCB : évolution de la moyenne de l'estimation du quantile 99 % en fonction de la taille du DOE. Les points représentent les évaluations de l'algorithme de la Figure 49. Chaque couleur correspond à une évaluation de l'algorithme. La courbe représente la pseudo-référence issue de l'algorithme en Figure 50. 85

Figure 54 : Modèle PCB : évolution de l'écart type de l'estimation du quantile 99 % en fonction de la taille du DOE. Les points représentent les évaluations de l'algorithme de la Figure 49. Chaque couleur correspond à une évaluation de l'algorithme. La courbe représente la pseudo-référence issue de l'algorithme en Figure 50. 85

Figure 55 : Géométrie de la cavité comportant deux ouvertures et une source de courant sur un fil conducteur à l'intérieur de la cavité. 86

Figure 56 : Histogramme de la sortie de l'échantillon (champ électrique maximum sur une sphère à 3 mètres de distance de la cavité). 88

Figure 57 : Estimation du quantile 99% de E par l'échantillon de référence et le SM en fonction du nombre d'appel au modèle HFSS i.e. le budget de simulation. 90

Figure 58 : Estimation du quantile 99% de E par la Stratification Contrôlée à la suite du SM en fonction du nombre d'appel au modèle HFSS i.e. le budget de simulation. 90

Figure 59 : Distribution des 16 variables d'entrées (bleu). Pour chaque variable, distribution des valeurs ayant provoquées des valeurs extrêmes de sortie (Orange). 92

Figure 60 : Géométrie de la cavité percée d'une ouverture et de la coque surmontée par l'antenne. 96

Figure 61 : Zoom de la Figure 60 sur l'antenne et la coque à l'intérieur de laquelle se trouve un équipement sensible (non figuré sur le schéma). 97

Figure 62 : Principe de normalisation des mesures de susceptibilité. 99

Figure 63 : p-valeurs du test statistique AD (IMV) à partir de l'échantillon de contrainte de référence pour plusieurs hypothèses de distribution. 100

Figure 64 : Histogramme de l'échantillon de référence de la contrainte (C) en Ampères, ainsi que les meilleurs ajustements pour chacune des lois non rejetées par le test d'AD (IMV). 100

Figure 65 : Densités de probabilité de la contrainte (C) et de la susceptibilité (S) de référence (Réf.) en Ampère et densité de probabilité de la défaillance (D) qui en découle. 101

2. Liste des Tableaux

Tableau 1: Principaux acronymes, et conventions de notations mathématiques utilisées. 9

Tableau 2: Variables spécifiques au chapitre II. 10

Tableau 3: Variables spécifiques au chapitre III. 10

Tableau 4: Variables spécifiques au chapitre IV. 10

Tableau 5 : Distributions a priori utilisées selon la loi sous-jacente supposée. 24

Tableau 6 : Estimateurs les plus probables lorsque la distribution est N avec une distribution a priori de $\boldsymbol{\theta} = [\mu, \sigma 2]$ uniforme (i.e. estimateur scalaire de l'IMV). Deux estimateurs sont possibles pour $\theta 2$: le premier ($\sigma n 2$) est biaisé alors que le deuxième ($\sigma n - 12$) ne l'est pas. 25

Tableau 7 : Estimateurs les plus probables lorsque la distribution est W avec une distribution a priori de $\boldsymbol{\theta} = [\gamma, \beta]$ uniforme (i.e. estimateur scalaire de l'IMV). 26

Tableau 8 : Proportion de prise de bonnes (vert)/mauvaises (rouge) décisions (%) concernant l'identification de la loi sous-jacente de seuil de susceptibilité pour l'équipement virtuel de la Figure 22. 42

Tableau 9 : Seuils de défaillance mesurés (Ampères) lors de tests d'injection de courant sur des alimentations (Famille A). 44

Tableau 10 : Seuils de défaillance mesurés (Ampères) lors de tests d'injection de courant sur des alimentations (Famille B). 45

Tableau 11 : Résultat de l'inférence du seuil de susceptibilité des deux familles d'alimentations. 45

Tableau 12 : Bilan des comparaisons entre l'IB et l'IMV dans chacun des cas traités dans ce chapitre, uniquement pour de petites tailles d'échantillon ($n \leq 40$). 47

Tableau 13 : Entrées du modèle PCB. 62

Tableau 14 : Entrées du modèle S11. 64

Tableau 15 : Temps de simulation consacré à la construction d'un seul KRI pour un seul DOE. 72

Tableau 16 : Estimation du nombre de réalisations nécessaires afin d'obtenir au moins n événements extrêmes en fonction de la probabilité p (%) d'événements pour un intervalle de confiance de 95 %. 80

Tableau 17 : Entrées du modèle du fil à l'intérieur de la cavité percée. 87

Tableau 18 : Entrées du modèle de l'antenne à l'intérieur de la cavité percée. 98

Tableau 19 : Probabilité de défaillance estimée (%), la référence étant de 0.13795 %, en fonction du nombre de réalisation utilisées pour estimer la contrainte (n_C) et celui pour la susceptibilité (n_S). 103

IV. Liste des publications

Revues à comité de lecture :

[1] T. Houret, P. Besnier, S. Vauchamp and P. Pouliguen, “Estimating the Probability Density Function of the Electromagnetic Susceptibility from a Small Sample of Equipment,” *Progress In Electromagnetics Research*, 2019.

[2] T. Houret, P. Besnier, S. Vauchamp and P. Pouliguen, “Kriging in addition to controlled Stratification: Practical Implementation for Extreme Quantile Estimation in EMC Risk Analysis,” *IEEE access*, 2019.

Conférences internationales à comité de lecture :

[1] T. Houret, P. Besnier, S. Vauchamp and P. Pouliguen, “Inférence Statistique du Niveau de Susceptibilité Electromagnétique à Partir de Tests Effectués sur Quelques Exemplaires d’un Equipement,” *Colloque International et Exposition sur la Compatibilité Electromagnétique*, Paris, France, 2018.

[2] T. Houret, P. Besnier, S. Vauchamp and P. Pouliguen, “Inferring the Probability Distribution of the Electromagnetic Susceptibility of Equipment from a Limited Set of Data,” *International Symposium and Exhibition on Electromagnetic Compatibility (EMC Europe)*, Amsterdam, Netherlands, 2018.

[3] T. Houret, P. Besnier, S. Vauchamp and P. Pouliguen, “Comparison of Surrogate Models for Extreme Quantile Estimation in the Context of EMC Risk Analysis,” *Joint International Symposium on Electromagnetic Compatibility and Asia-Pacific International Symposium on Electromagnetic Compatibility (EMC/APEMC)*, Sapporo, Japan, 2019.

[4] T. Houret, P. Besnier, S. Vauchamp and P. Pouliguen, “Combining Kriging and Controlled Stratification to Identify Extreme Levels of Electromagnetic Interference,” *International Symposium and Exhibition on Electromagnetic Compatibility (EMC Europe)*, Barcelona, Spain, 2019.

[5] T. Houret, P. Besnier, S. Vauchamp and P. Pouliguen, “Estimation de la probabilité de défaillance basée sur la méthode Krigeage-Stratification Contrôlée pour l’analyse de contraintes extrêmes et sur l’inférence statistique de la susceptibilité,” *Colloque International et Exposition sur la Compatibilité Electromagnétique*, Lyon, France, 2020 (submitted).

[6] T. Houret, P. Besnier, S. Vauchamp and P. Pouliguen, “Probability of Failure Using the Kriging - Controlled Stratification Method and Statistical Inference,” *International Symposium and Exhibition on Electromagnetic Compatibility (EMC Europe)*, Rome, Italy, 2020 (submitted).

Workshop :

[1] T. Houret, P. Besnier, S. Vauchamp and P. Pouliguen, “Inferring the Probability Distribution of the Electromagnetic Susceptibility of Equipment from a Limited Set of Data,” *UMEMA*, Torino, Italy, 2017.

V. Table des matières

Remerciements 3

Résumé 5

Abstract 6

Sommaire 7

Chapitre I. Introduction générale 11

I. Contexte 11
II. Problématique et objectif 12
III. Positionnement de la thèse 13
IV. Organisation du manuscrit 15

Chapitre II. Analyse probabiliste de la susceptibilité d'équipements électroniques 17

I. Introduction 17
II. Etat de l'art 17
1. Démarche expérimentale des mesures de susceptibilité 17
2. Traitement des données mesurées par inférence statistique 19
a. Rappels sur les distributions et leurs paramètres 19
b. Inférence paramétrique 22
i. Densité a priori 23
ii. Fonction de vraisemblance 24
iii. Encadrement probabiliste de la probabilité de défaillance 26
III. Estimation du type de distribution 27
1. Coefficient de variation et proximité entre distributions 27
2. Tests statistiques 28
3. Test du BIC 29
4. Comparaison AD/BIC 30
IV. Estimation des paramètres de la distribution 31
1. Cas de la distribution Normale 32
2. Cas de la distribution de Weibull 33
3. Conclusion sur l'estimation des paramètres 34
V. Encadrement probabiliste du seuil de susceptibilité 34
1. Validation de l'encadrement 34
2. Encadrement dans le cas d'une distribution Normale 37
3. Encadrement dans le cas d'une distribution Weibull 38
VI. Applications CEM 39
1. Etude de la susceptibilité d'un équipement virtuel 39

a. Présentation du cas d'étude 39
b. Identification de la loi 41
c. Encadrement probabiliste du seuil de susceptibilité 42

2. Etude de la susceptibilité d'alimentations 44

a. Présentation du cas d'étude 44
b. Encadrement probabiliste du seuil de susceptibilité 45

VII. Conclusion et pistes de réflexion 46

1. Conclusion 46

2. Pistes de réflexion 47

Chapitre III. Recherche de valeurs extrêmes de contrainte par krigeage et stratification contrôlée 49

I. Introduction 49

II. Etat de l'art 50

1. Estimation empirique d'un quantile par simulation de Monte Carlo (MCS) 51

2. Estimation d'un quantile par métamodèle 52

a. Définition et métamodèles choisis 52
b. Métamodèle pour la recherche d'événements extrêmes 53
c. Comparaisons ente les métamodèles en CEM 53

3. Support Vector Machine (SVM) 53

4. Réseau de Neurones (NN) 55

5. Polynomial chaos expansion (PCE) 56

6. Krigeage (KRI) 57

7. Le polynomial chaos expansion associé au krigeage (PCE_KRI) 59

8. Stratification Contrôlée (SC) 59

III. Modèles utilisés 61

1. Modèle PCB 61

2. Modèle S_{11} 63

IV. Comparaison des performances des métamodèles 65

1. Critères de performance 66

2. Modèle PCB : Résultats 66

3. Modèle S11 : Résultats 68

4. Sélection d'un métamodèle : Le krigeage 69

V. Intérêt de la Stratification Contrôlée par rapport au krigeage seul 70

1. Algorithme de comparaison et critères de performance 70

2. Modèle S11 : Méthodologie et résultats 71

3. Modèle PCB : Méthodologie et résultats 73

4. Conclusion sur l'intérêt de la stratification contrôlée 76

VI. Gestion du budget de simulation lors de l'estimation d'un quantile extrême par la méthode K-SC 77

1. Présentation de l'algorithme de gestion du budget de simulation 77
a. Description de l'algorithme 78
b. Gestion du DOE de taille maximale initiale Smax (section rouge de la Figure 49) 79
c. Construction de différents SMs (krigeage) et simulations Monte Carlo associées (section bleue de la Figure 49) 80
d. Stratification contrôlée (section violette de la Figure 49) 81
2. Validation de l'algorithme sur les modèles analytiques 81
a. Description de la méthodologie 81
b. Modèles PCB et S11 : Résultats 82
i. Résultats de validation avec le modèle S11 82
ii. Résultats de validation avec le modèle PCB 84
3. Cas d'application 85
a. Présentation du modèle d'étude 86
b. Echantillon de référence 87
c. Résultats 88
i. Choix du DOE initial 88
ii. Simulation Monte Carlo du SM 88
iii. Enrichissement du DOE 88
iv. Estimation du quantile par Stratification Contrôlée 89
d. Analyse des valeurs extrêmes obtenues 90
4. Bilan sur l'intérêt pratique de cette implémentation de la Stratification Contrôlée 92
VII. Conclusion et pistes de réflexion 93
1. Conclusion 93
2. Pistes de réflexion 93

Chapitre IV. Estimation de la probabilité de défaillance d'équipements électroniques en cas d'agression électromagnétique 95

I. Introduction 95
II. Présentation du cas d'étude 96
1. Equipement soumis à une agression électromagnétique et chemin de couplage associé 96
2. Mise en cohérence des niveaux de contrainte et de susceptibilité 98
III. Obtention d'une probabilité de défaillance de référence 99
1. Contrainte de référence 99
2. Susceptibilité de référence 100
3. Probabilité de défaillance de référence 100
IV. Estimation de la probabilité de défaillance 101
1. Nouvelle stratégie de stratification pour la SC 102
2. Résultats de l'estimation de la probabilité de défaillance 102
V. Conclusion et piste de réflexion 103

Chapitre V. Conclusion générale 105

I. Rappel de l'objectif et de la stratégie 105

II. Synthèse de l'étude de la susceptibilité 105
III. Synthèse de l'étude de la contrainte 106
IV. Synthèse de l'étude de la probabilité de défaillance 106

Références 107

Annexes 115

I. Annexes du Chapitre II 115
1. Inférence par méthode des moments 115
2. Implémentation de l'inférence bayésienne en Matlab 116
II. Annexes du Chapitre III 117
1. Calcul du modèle PCB 117
2. Implémentation de la stratification contrôlée en Matlab 118
III. Liste des Figures et des Tableaux 122
1. Liste des Figures 122
2. Liste des Tableaux 124
IV. Liste des publications 126
V. Table des matières 127

INSA de RENNES
Service des Formations

AVIS DU JURY SUR LA REPRODUCTION DE LA THESE SOUTENUE

Titre de la thèse:
Méthode d'estimation des valeurs extrêmes des distributions de contraintes induites et de seuils de susceptibilité dans le cadre des études de durcissement et de vulnérabilité aux menaces électromagnétiques intentionnelles

Nom Prénom de l'auteur : HOURET THOMAS

Membres du jury :
- Monsieur PICHON Lionel
- Monsieur LALLECHERE Sébastien
- Monsieur POULIGUEN Philippe
- Monsieur VAUCHAMP Stéphane
- Monsieur BESNIER Philippe
- Monsieur REINEIX Alain
- Madame RICHALOT Elodie

Président du jury : PICHON Lionel

Date de la soutenance : 19 Décembre 2019

Reproduction de la these soutenue

☒ Thèse pouvant être reproduite en l'état
☐ Thèse pouvant être reproduite après corrections suggérées

Fait à Rennes, le 19 Décembre 2019

Signature du président de jury

Le Directeur,

M'hamed DRISSI

UNIVERSITE BRETAGNE LOIRE / MATHSTIC

INSA RENNES

Titre : Méthode d'estimation des valeurs extrêmes des distributions de contraintes induites et de seuils de susceptibilité dans le cadre des études de durcissement et de vulnérabilité aux menaces électromagnétiques intentionnelles

Mots clés : IEMI, Susceptibilité électromagnétique, Inférence statistique, Propagation d'incertitudes, Stratification contrôlée, Krigeage

Résumé : Une Interférence Electromagnétique Intentionnelle (IEMI) agressant un équipement électronique peut provoquer une défaillance de ce dernier. L'étude des effets d'une IEMI commence par l'évaluation du risque de défaillance afin de mettre en place les protections adéquates. Malheureusement, une prédiction déterministe d'une défaillance est impossible car les caractéristiques de l'équipement et de l'agression sont très incertaines. La stratégie proposée consiste à modéliser la contrainte générée par l'agression, ainsi que la susceptibilité de l'équipement, comme des variables aléatoires. Ensuite, trois étapes sont nécessaires. La première concerne l'estimation de la distribution probabiliste de la variable aléatoire de susceptibilité, la seconde porte sur celle de la contrainte, pour enfin, dans une troisième étape, en déduire la probabilité de défaillance.

Pour la première étape, nous utilisons des méthodes d'inférence statistique sur un petit échantillon de seuils de susceptibilités mesurés. Nous comparons deux types d'inférence paramétrique : bayésienne et celle du maximum de vraisemblance. Nous concluons qu'une approche pertinente pour l'analyse du risque CEM consiste à utiliser les intervalles de confiance ou de crédibilité des estimations des paramètres pour encadrer la probabilité de défaillance, quelle que soit la méthode d'inférence choisie.

Pour la deuxième étape, nous explorons les techniques de recherche de valeurs extrêmes tout en réduisant le nombre de simulations nécessaires. En particulier, nous proposons la technique de la stratification contrôlée par un métamodèle de krigeage. Nous montrons que cette technique améliore drastiquement les performances par rapport à l'approche classique (simulation Monte Carlo). De plus, nous proposons une implémentation particulière de cette technique afin de maitriser le coût de calcul.

Enfin, la troisième étape est la plus simple une fois les deux premières franchies puisque, par définition, une défaillance survient lorsque la contrainte est supérieure à la susceptibilité. A partir d'un cas test final comportant la simulation de l'agression d'un équipement et de données de susceptibilité de cet équipement, nous calculons un encadrement de la probabilité de défaillance en recourant aux méthodes développées pendant cette thèse. En particulier, nous montrons que l'utilisation conjointe de la stratification contrôlée par le krigeage et de l'inférence de la distribution de la susceptibilité permet effectivement d'encadrer l'estimation de la vraie probabilité de défaillance.

Title : Method for estimating extreme values of induced stress distributions and susceptibility thresholds in studies of hardening and vulnerability to intentional electromagnetic threats

Keywords : IEMI, Electromagnetic susceptibility, Statistical inference, Uncertainty propagation, Controlled Stratification, Kriging

Abstract: Intentional ElectroMagnetic Interference (IEMI) can cause equipment failure. The study of the effects of an IEMI begins with an assessment of the risk of equipment failure in order to implement appropriate protections, if required. Unfortunately, a deterministic prediction of a failure is impossible because both characteristics of equipment and of the aggression are very uncertain. The proposed strategy consists of modelling the stress generated by the aggression, as well as the susceptibility of the equipment, as random variables. Then, three steps are necessary: The first step deals with the estimation of the probability distribution of the random susceptibility variable.
The second step deals with the similar estimation for the constraint / stress then that of the stress. Eventually, the third step concerns the calculation of the probability of failure.

For the first step, we use statistical inference methods on a small sample of measured susceptibility thresholds. We compare two types of parametric inference: bayesian and maximum likelihood. We conclude that a relevant approach for a risk analysis is to use the confidence or credibility intervals of parameter estimates to frame the probability of failure, regardless of the inference method chosen.

For the second step we explore extreme value exploration techniques while reducing the number of simulations required. In particular, we propose the technique of Controlled Stratification by Kriging. We show that this technique drastically improves performance compared to the classic approach (Monte Carlo simulation). In addition, we propose a particular implementation of this technique in order to control the calculation effort.

Finally, the third step is the simplest once the first two steps have been completed since, by definition, a failure occurs when the stress is greater than the susceptibility. With the help of a final test case comprising the simulation of an electromagnetic aggression on a piece of equipment, we use the method developed in our work to estimate the frame of the probability of failure, More specifically, we show that the combined use of controlled stratification by kriging and inference of susceptibility distribution, allows to frame the estimated true value of the probability of failure.