



HAL
open science

Développement d'un dialogue essai-calcul dans un cadre probabiliste pour la validation de châssis automobiles

Han Guo

► **To cite this version:**

Han Guo. Développement d'un dialogue essai-calcul dans un cadre probabiliste pour la validation de châssis automobiles. Mécanique [physics.med-ph]. Université de Technologie de Compiègne, 2022. Français. NNT : 2022COMP2671 . tel-03970187

HAL Id: tel-03970187

<https://theses.hal.science/tel-03970187v1>

Submitted on 2 Feb 2023

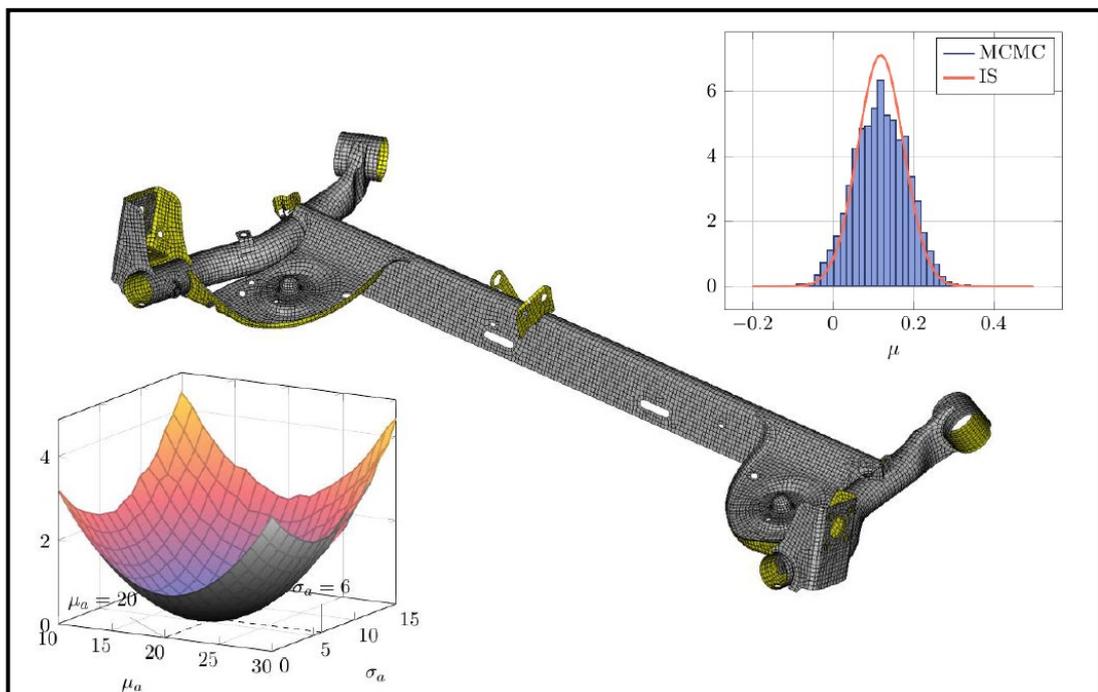
HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

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.

Par Han GUO

*Développement d'un dialogue essai-calcul
dans un cadre probabiliste pour la validation
de châssis automobiles*

Thèse présentée
pour l'obtention du grade
de Docteur de l'UTC



Soutenue le 16 février 2022

Spécialité : Mécanique Numérique : Unité de recherche en
Mécanique - Laboratoire Roberval (FRE UTC - CNRS 2012)

D2671

UNIVERSITÉ DE TECHNOLOGIE DE COMPIÈGNE
ÉCOLE DOCTORALE ED71
Sciences pour l'ingénieur

T H È S E

**Développement d'un dialogue essai-calcul dans
un cadre probabiliste pour la validation de
châssis automobiles**

Spécialité : Mécanique Numérique

Présentée par Han GUO

16 Février 2022

Jury composé de :

Guillaume PUEL, CentraleSupélec, MSSMat (rapporteur)
Vlad Stefan BARBU, Université de Rouen Normandie, LMRS (rapporteur)
Frédéric DRUESNE, UTC, Roberval (directeur)
Nikolaos LIMNIOS, UTC, LMAC (directeur)
Pierre FEISSEL, UTC, Roberval
Salim BOUZEBDA, UTC, LMAC
Pascal LARDEUR, UTC, Roberval (président)
Nadia ACHOUR-RENAULT, Renault
Alain PATIGNIEZ, Renault
Stéphane BOUYAUX, Renault

致亲爱的父母。
For my beloved parents.

Remerciement

Je voudrais ici remercier toutes les personnes qui ont rendu ce travail de thèse possible, mes excuses à l'avance à tous ceux que je ne vais pas citer mais qui ont compté pour moi.

Tout d'abord, mes remerciements s'adressent à Alain Patigniez et Stéphane Bouyaux pour leur accueil au sein respectivement des groupes déploiement et simulation numérique des essais de validation de châssis de la Direction Plateforme chez Renault sur le site du Mans. Ils m'ont alimenté en données et en moyens les parties expérimentale et numérique de la thèse, ils ont également suivi l'avancée des travaux pendant ces années de thèse.

Je voudrais remercier Guillaume Puel et Vlad Stepfan Barbu pour avoir évalué la qualité de ce manuscrit, ainsi que pour leurs suggestions constructives. Je tiens à remercier Pascal Lardeur pour avoir présidé le jury, ainsi que pour l'intérêt qu'il a manifesté durant la thèse en tant que membre du comité de suivi individuel avec Zohra Cherfi-Boulangier. Je voudrais également remercier Nadia Achor-Renault pour avoir participé à ce jury. Je tiens à exprimer ma gratitude à tous les membres du jury pour avoir assisté à la soutenance de ma thèse.

J'adresse mes remerciements à mon directeur de thèse Frédéric Druesne, pour sa patience, sa disponibilité et ses conseils, ainsi que pour son aide précieuse dans la rédaction du manuscrit et d'un article.

Je tiens à remercier mon directeur de thèse Nikolaos Limnios pour ses enseignements, sa disponibilité, sa responsabilité, ses encouragements, ainsi que pour les discussions multidisciplinaires en mathématique et en mécanique qui m'ont permis de découvrir un autre monde.

Je souhaite témoigner ma reconnaissance à mon encadrant de thèse Pierre Feissel, pour ses contributions ainsi que pour les discussions qui m'ont guidé dans ma réflexion.

Je voudrais remercier mon encadrant de thèse Salim Bouzebda pour ses explications pédagogiques et ses encouragements durant la thèse.

Je tiens à remercier les collègues chez Renault, Dimitri, Cédric, Yannick, Céline, Thierry et Aude, pour leur aide à la compréhension du contexte industriel et pour la collecte des données des essais. Je remercie également les stagiaires Julien, Jason et Hugo pour l'exploitation des travaux de thèse sur plusieurs aspects au sein du groupe Renault et pour avoir fourni des données de simulation qui ont alimenté la partie application du manuscrit.

Je remercie également tous mes collègues du laboratoire Roberval et du laboratoire de mathématiques appliquées et mes amis. Merci à Sini et Meihua pour leurs accompagnements et leur soutien. J'adresse mes remerciements spéciale à Ines pour sa gentillesse

et les discussions sur plusieurs aspects. Je tiens à remercier Aurélien pour son accompagnement et ses aides notamment sur Latex pendant la période de rédaction. Je voudrais remercier Lia (Dongliang) et Linhui pour les partages de vie et leurs encouragements. Merci spécialement à Vincent qui a partagé le bureau avec moi pendant trois ans, qui m'a aidé à m'intégrer dans un environnement francophone. Merci à Monsieur Duchemin, professeur d'alto au conservatoire de Compiègne, qui m'influence avec ses passions de la musique et de la vie. J'envoie tous mes remerciements à Ye pour ses soins et les discussions qui m'ont beaucoup aidé.

Je m'adresse particulièrement mes remerciements à tous les amis, anciens et nouveaux de l'UTC : Jing, Guoqiang, Jishuai, Jeanne, Loan, Daoming, Jichang, Jiashuo, Yunbo, Dali, Jolanthe, Cédric, Yoanne, Nesrine, Aurélien, Kwatar, Katherine, Clarissa, Ludovic, Caroline, Jianqiang, Xin, Ye, Yuqi, Junting, Mélissa, Nicolas, Joey, Dong, Zhaoyuan, Xinyu, Bo, Zhongtian, Xiangjun, Uy, Phong, Jaoued, Inass, Wang, Zhao, Qing, Ana, et aussi les amis lointains : Jiaxu et Yakun.

J'adresse toute ma reconnaissance à mes parents pour leur soutien, amour et compréhension, vous êtes toujours derrière de moi, vous me donnez du courage.

Acknowledgement

I would like to thank all the people who made this thesis possible, and I apologise in advance to all those who I will not mention but who meant a lot to me.

First of all, I would like to acknowledge Alain Patigniez and Stéphane Bouyaux for their welcome in the deployment and numerical simulation groups of the chassis validation tests of the Platform Department at Renault in Le Mans. They provided me with data and tools for the experimental and numerical parts of the thesis, they also followed the progress of the work during these years.

I would like to thank Guillaume Puel and Vlad Stepfan Barbu for evaluating the quality of this manuscript and for their constructive suggestions. I would like to thank Pascal Lardeur for chairing the jury, as well as for the interest he showed during the thesis as a member of the individual monitoring committee with Zohra Cherfi-Boulanger. I also want to thank Nadia Achor-Renault for participating in the jury. One more time, I want to express my gratitude to all the members of the jury for attending the defence of my thesis.

I would like to thank my thesis supervisor Frédéric Druesne, for his patience, availability and advice, as well as for his precious help in writing the manuscript and an article.

I wish to show my appreciation to my thesis supervisor Nikolaos Limnios for his teaching, his availability, his responsibility, his encouragement, as well as for the multi-disciplinary discussions in mathematics and mechanics which allowed me to discover another world.

I want to express my gratitude to my thesis supervisor Pierre Feissel, for his contributions as well as for the discussions that guided me in my reflection.

I would like to acknowledge my thesis supervisor Salim Bouzebda for his pedagogical explanations and encouragement during the thesis.

I would like to thank my colleagues at Renault, Dimitri, Cédric, Yannick, Céline, Thierry and Aude, for their help in understanding the industrial context and for collecting test data. I would like also thank the internship students Julien, Jason and Hugo for the exploitation of the thesis work on several aspects within the Renault group and for providing simulation data that supplied the application part of the manuscript.

I want to thank all my colleagues from the laboratories of Roberval and the Applied Mathematics and all of my friends. Thanks to Sini and Meihua for their accompany and support. Special thanks go to Ines for her kindness and the discussions in several aspects. Thanks to Aurélien for his encouragements especially his help on Latex during the writing period. I would like to thank Lia (Dongliang) and Linhui for sharing and their

encouragements. Thanks to Vincent who shared the office with me since three years, who helped me to integrate into a French speaking environment. Thanks to Mr. Duchemin, viola teacher at the Conservatoire de Compiègne, who influences me with his passion for music and life. Thanks to Ye for his concern and discussions which helped me a lot.

I wish to extend my special thanks to all my friends with whom I have shared my life in UTC : Jing, Guoqiang, Jishuai, Jeanne, Loan, Daoming, Jichang, Jiashuo, Yunbo, Dali, Jolanthe, Cédric, Yoanne, Nesrine, Aurélien, Kwatar, Katherine, Clarissa, Ludovic, Caroline, Jianqiang, Xin, Ye, Yuqi, Junting, Mélissa, Nicolas, Joey, Dong, Zhaoyuan, Xinyu, Bo, Zhongtian, Xiangjun, Uy, Phong, Jaoued, Inass, Wang, Zhao, Qing, Ana, et aussi les amis lointains : Jiaxu et Yakun.

I am very grateful to my parents for their support, love and understanding, you are always behind me and give me courage.

Résumé

La fiabilité est une dimension essentielle pour la conception d'un véhicule dans l'industrie automobile. Pour le groupe Renault, la fiabilité des composants d'un châssis est validée à l'aide d'essais accélérés de fatigue, accompagnés de simulation numérique. Ces essais expérimentaux mettent en évidence une dispersion sur la durée de vie des structures. Afin d'améliorer le processus de validation, il est visé de prendre en compte les incertitudes dans la modélisation numérique pour que la variabilité de la durée de vie reproduise celle observée lors des essais. Avec l'hypothèse que les incertitudes sont uniquement liées à la courbe de Wöhler, un problème d'identification du modèle probabiliste de fatigue a été construit. Une méthodologie d'identification composée de deux étapes a été proposée.

Tout d'abord, la durée de vie expérimentale sous forme de données groupées, censurées par intervalle, a été estimée avec la méthode du maximum de vraisemblance et la méthode Bayésienne pour caractériser sa variabilité. Trois lois adaptées aux données ont été utilisées. Une étude de la fiabilité a également été menée.

Puis, l'identification a été réalisée par une approche analytique et une approche simulation. Dans le cadre de l'approche simulation, différentes fonctions coûts ont été proposées et optimisées pour déterminer les paramètres de la courbe de Wöhler. Une étude sur la précision d'identification du modèle liée à la censure des données groupées et sur l'erreur liée à la méthode d'estimation a été réalisée à l'aide d'un réseau à fonctions de base radiale.

La méthodologie d'identification a été validée sur deux projets industriels. Une interface a également été développée sous MATLAB pour faciliter l'utilisation de cette approche.

Abstract

Reliability is essential to be considered for designing a car in automotive industry. For Renault company, the reliability of chassis components is validated by accelerated fatigue tests associated with numerical simulations. These experimental tests highlight dispersion for fatigue life of structures. To improve the validation process, it is aimed to take into account uncertainties in the numerical modeling so that the variability of the estimated fatigue life reproduces that observed during the tests. Assuming that all uncertainties are included in the S-N curve, an identification problem of a probabilistic fatigue model has been built. An identification strategy composed by two steps has been proposed.

Firstly, the experimental fatigue lifetime described by grouped data (interval-censored) have been estimated by maximum likelihood and Bayesian methods to characterize its variability. Three laws adapted to the data were used. A study of the reliability has also been investigated.

Secondly, the identification was done by analytical approach and simulation approach. For the simulation approach, different cost functions were proposed and optimised by the optimisation algorithms to determine the parameters of the S-N curve. The accuracy of the identification was also studied, a radial basis function network was used instead of the optimisation of the cost functions.

The methodology has been validated with two industrial projects. An interface has been developed in MATLAB to simplify the use of the identification methodology.

Table des matières

Table des matières	i
Table des figures	v
Liste des tableaux	ix
Nomenclature	xi
Introduction générale	1
1 État de l'art sur la fatigue des structures automobiles	5
1.1 Validation en fiabilité et tests d'endurance	6
1.1.1 Objectif de fiabilité	6
1.1.2 Cycle en V de la validation	7
1.1.3 Trois essais d'endurance	8
1.1.4 Liaison au sol	10
1.1.5 Essai synthétique sur essieu souple et mesure de durée de vie	12
1.1.6 Simulation numérique de l'essai synthétique	13
1.2 Modélisation de la fatigue	14
1.2.1 Calcul du dommage sous chargement avec amplitude variable	14
1.2.2 Modèle déterministe de la courbe de Wöhler	18
1.2.3 Modèle probabiliste de la courbe de Wöhler	19
1.3 Conclusion	20
2 Estimation of grouped data	23
2.1 Introduction	25
2.2 Presentation of experimental grouped data	26
2.2.1 Homogeneous grouped data	26
2.2.2 Non-homogeneous grouped data	26
2.3 MLE for Weibull and Log-normal distributions	27
2.3.1 Method and analysis	27
2.3.2 Weibull distribution	31
2.3.3 Log-normal distribution	34
2.3.4 Comparison of estimations obtained by Weibull and Log-normal distributions	36
2.4 Goodness of fit tests for grouped data	39

2.4.1	Kolmogorov-Smirnov test for grouped data	39
2.5	Bayesian estimation for Weibull distribution and Log-normal distribution	43
2.5.1	Method of Bayesian estimation	44
2.5.2	Bayesian estimation for Weibull distribution	47
2.5.3	Bayesian estimation for Log-normal distribution	51
2.5.4	Comparison among the different Bayesian estimation methods	55
2.6	Estimation with multinomial distribution for grouped data	57
2.6.1	Definition	57
2.6.2	Conjugate prior and posterior distribution	58
2.7	Simulated data estimation results	60
2.7.1	Simulating method and simulated samples	61
2.7.2	Bayesian estimation for Log-normal distribution with grouped data	63
2.8	Conclusion	65
3	Stratégie d'identification des incertitudes liées à la courbe de Wöhler	67
3.1	Introduction	69
3.2	Problématique	70
3.3	Approche analytique	70
3.3.1	Propagation directe/inverse avec un paramètre stochastique	71
3.3.2	Propagation directe avec deux paramètres stochastiques	71
3.4	Approche par simulation	72
3.4.1	Principe de l'approche par simulation	72
3.4.2	Construction des fonctions de coût	73
3.4.3	Méthode d'optimisation	77
3.5	Cas d'application - identification des deux paramètres de la courbe de Wöhler probabiliste	78
3.5.1	Données de référence	79
3.5.2	Identification des deux paramètres de la courbe de Wöhler P-S-N	79
3.5.3	Influence de la taille de l'échantillon sur la robustesse des fonctions de coût et de l'algorithme d'optimisation	81
3.5.4	Performance des fonctions de coût au voisinage du point de référence	86
3.6	Cas d'application - identification des trois paramètres de la courbe P-S-N	93
3.6.1	Identification à partir des données d'un seul cordon de soudure	93
3.6.2	Identification à partir des données de plusieurs cordons de soudure	95
3.6.3	Stratégie d'optimisation multi-niveaux adaptée	98
3.6.4	Identification des trois paramètres pour une autre référence	99
3.7	Précision des méthodologies d'identification	99
3.7.1	Génération des échantillons de μ_t et σ_t	102
3.7.2	Réseau de fonctions à base radiale	105
3.7.3	Résultat de l'étude sur la précision des méthodologies d'identification	109
3.8	Conclusion	113

4	Application industrielle	115
4.1	Introduction	116
4.2	Identification des paramètres probabilistes de la courbe de Wöhler pour deux essieux	116
4.2.1	Présentation des essieux X52 et X98	116
4.2.2	Résultats d'identification des paramètres probabilistes de la courbe de Wöhler	118
4.3	Exploitation des résultats de la thèse	122
4.3.1	Méthode d'estimation	122
4.3.2	Propagation des incertitudes sur la durée de vie	122
4.3.3	Critère de corrélation	124
4.4	Conclusions et perspectives	125
	Conclusion générale et perspectives	127
	Annexes	131
A	Calculation of Fisher information for grouped data	131
A.1	Fisher information of Weibull distribution	131
A.2	Fisher information of Log-normal distribution	132
B	Non-homogeneous grouped data	135
C	Bayes estimation for Log-normal distribution of exact data with conjugate prior distribution	137
D	Calcul de la statistique de test de Cramér-von Mises	139
E	Vérification de la densité de probabilité	141
F	Interface d'identification de la variabilité liée à la courbe de Whöler	143
	Bibliographie	147

Table des figures

1.1	Relation entre l'objectif de validation du profil de mission et les tests sur système ou composant	7
1.2	Cycle en V pour la validation du déploiement	8
1.3	Essai de roulage	9
1.4	Essai dynamique	9
1.5	Essieu instrumenté avant l'essai et essai synthétique	10
1.6	Système liaison au sol (Clio 4 2019)	11
1.7	Trains roulants	11
1.8	Essieu souple automobile	12
1.9	Fissure proche d'un cordon de soudure identifiée par ressuage	12
1.10	Modèle CAO et maillage éléments finis d'un essieu	13
1.11	Courbe de Wöhler	15
1.12	Superposition de la moyenne et amplitude de contrainte	16
1.13	Influence de la valeur de contrainte moyenne du chargement cyclique sur la courbe de Wöhler	16
1.14	Corrections de Goodman et Gerber	17
1.15	Courbe de Wöhler représentée sur une échelle logarithmique	19
1.16	Courbe de Wöhler probabiliste (courbe P-S-N) représentée sur une échelle logarithmique	20
1.17	Dialogue essai-calcul	21
2.1	Grouped data	26
2.2	Non-homogeneous grouped data	27
2.3	Example of Renault grouped data	32
2.4	PDF of the Weibull ($\hat{\alpha} = 1.2832, \hat{\beta} = 3.6875$) and Log-normal ($\hat{\mu} = 0.1178, \hat{\sigma} = 0.2886$) distributions	36
2.5	Reliability analysis with Log-normal and Weibull distributions for homogeneous grouped data	37
2.6	PDF of Weibull ($\hat{\alpha} = 2.6651, \hat{\beta} = 1.0733$) and Log-normal distributions ($\hat{\mu} = 0.6171, \hat{\sigma} = 1.2965$)	38
2.7	Reliability analysis with Log-normal and Weibull distributions for non-homogeneous data	38
2.8	Schematic of the KS test for the grouped data	40
2.9	Theoretical and empirical cumulative distributions of simulated grouped data sample from Log-normal distribution	42

2.10	Theoretical and empirical cumulative distributions of experimental grouped data for Log-normal and Weibull distributions	43
2.11	PDF estimated by MLE and Bayes methods for Log-normal distribution	49
2.12	Normalized histogram of posterior parameters	51
2.13	Trajectory of MCMC sampling	51
2.14	Normalized histogram of posterior sample of μ by MCMC	52
2.15	Posterior parameter μ sampled by MH and calculated by importance sampling	53
2.16	Histogram of posterior distribution of parameters with normal-inverse gamma prior	55
2.17	Histogram of posterior distribution of parameters with uniform prior	56
2.18	Comparison of PDF estimated by MLE with different estimation methods for Weibull and Log-normal distributions	56
2.19	Comparison of reliability function and failure rate obtained by MLE and different estimation methods for Weibull and Log-normal distributions	57
2.20	Comparison of reliability obtained by MLE estimation of Weibull and Log-normal distributions with that obtained by Bayesian estimation of multinomial distribution	60
2.21	Comparison of reliability obtained by MLE estimation of Weibull and Log-normal distributions with that obtained by Bayesian estimation of multinomial distribution with non-informative prior	61
2.22	Estimated Log-normal parameters with uncensored and grouped data	62
2.23	Log-normal density distribution estimated by MLE from simulated uncensored and grouped data	62
2.24	Log-normal density distribution estimated by MLE from simulated uncensored and grouped data	63
2.25	Histogram of posterior Log-normal parameters with Normal-gamma prior for simulated data	64
2.26	Histogram of posterior Log-normal parameters with uniform prior for simulated data	64
2.27	Reliability estimated by different methods	65
3.1	Schéma de l'algorithme par essais particuliers	78
3.2	Valeurs des six fonctions de coût dans l'espace des paramètres (μ_a, σ_a)	80
3.3	Médiane, 10 ^{ème} et 90 ^{ème} percentiles des six fonctions de coût pour différentes tailles d'échantillon T^{EF}	82
3.4	Histogramme des fonctions de coût CF2, CF3 et CF4 pour 400 répétitions	84
3.5	Rapport entre la longueur de l'intervalle entre 10 ^{ème} et 90 ^{ème} percentiles pour différentes tailles d'échantillon et la valeur moyenne de la fonction de coût dans la zone $\pm 5\%$	85
3.6	Optimisation des paramètres μ_a et σ_a par l'algorithme à essaim de particules pour les fonctions coût CF1, CF3 et CF4	87
3.7	Optimisation des paramètres μ_a et σ_a par l'algorithme à essaim de particules pour les fonctions coût CF5 et CF6	88

3.8	Médianes, 10 ^{ème} et 90 ^{ème} percentiles des six fonctions de coût avec $\sigma_a = 6$ pour 400 répétitions	90
3.9	Médianes, 10 ^{ème} et 90 ^{ème} percentiles des six fonctions de coût avec $\mu_a = 20$ pour 400 répétitions	91
3.10	La médiane, 10 ^{ème} et 90 ^{ème} percentiles des fonctions de coût pour 400 répétitions	92
3.11	Médiane, 10 ^{ème} et 90 ^{ème} percentiles des six fonctions de coût pour différentes tailles d'échantillon T^{EF}	94
3.12	Fonctions de coût CF1 et CF4 dans l'espace des paramètres pour σ_a fixé	95
3.13	Fonction de coût CF4 dans l'espace paramétrique pour σ_a fixé avec multi joints de soudure	96
3.14	Groupe de la courbe S-N optimisée	96
3.15	Groupe de la courbe S-N optimisée en échelle de logarithme décimal	97
3.16	Fonctions de coût dans l'espace des paramètres avec σ_a fixé	97
3.17	Valeur des six fonctions de coût dans l'espace des paramètres (μ_a, σ_a) avec la seconde référence	100
3.18	Courbes S-N optimisées et données de 1 à 6	101
3.19	Échantillon uniforme généré par la séquence de Halton et par la méthode d'hypercube latin	103
3.20	Échantillon Normal généré par la séquence de Halton et la méthode d'hypercube latin	104
3.21	Échantillons de μ_t et σ_t par deux approches	105
3.22	Schéma du réseau de fonctions à base radiale	106
3.23	Entraînement du RBF pour (a) μ_a et (c) σ_a . Validation du RBF pour (b) μ_a et (d) σ_a avec la fonction de coût CF1	109
3.24	Entraînement du RBF pour la fonction de coût CF1	110
3.25	Entraînement du RBF pour (a) μ_a et (c) σ_a . Validation du RBF pour (b) μ_a et (d) σ_a avec la fonction de coût CF3	111
3.26	Entraînement du RBF pour (a) μ_a et (c) σ_a . Validation du RBF pour (b) μ_a et (d) σ_a avec la fonction de coût CF4	112
3.27	Paramètres μ_a et σ_a prédits par le réseau RBF de la fonction coût CF1	113
4.1	Essieu arrière X52	117
4.2	Essieu X52 et emplacements des paires de cordons de soudure étudiés	117
4.3	Essieu arrière X98	118
4.4	Essieu X98 et emplacement des paires de cordons de soudure étudiés	118
4.5	Application de la méthode d'estimation du maximum de vraisemblance pour des durées de vie de type données groupées expérimentales	123
4.6	Propagation de l'incertitude de la courbe de Wöhler sur la durée de vie	123
4.7	Intervalle de confiance à 90% pour le critère de corrélation	124
E.1	Densité de probabilité du T	142
F.1	Feuille Excel des données expérimentales	144
F.2	Etape 1 de l'interface	144

TABLE DES FIGURES

F.3	Paramètres estimés à partir des données expérimentales	145
F.4	Etape 2 de l'interface	145
F.5	Fichier de résultat	146

Liste des tableaux

2.1	Estimator $\tilde{\gamma}$ calculated by the approximation and importance sampling methods	49
2.2	Estimates of α and β under squared error loss function (mean) and absolute value loss function (median) with MCMC method	50
3.1	Fonctions coût proposées pour les méthodologies d'identification	76
3.2	Résultats d'optimisation avec les méthodes RS, AG et EP	77
3.3	Erreur relative sur les paramètres identifiés pour les fonctions de coût avec une taille d'échantillon égale à 2000	86
3.4	Optimisation en deux temps de trois paramètres pour différentes fonctions de coût	95
3.5	Optimisation des trois paramètres avec différentes fonctions de coût	98
3.6	Optimisation des 3 paramètres en deux étapes pour les différentes fonctions coût	98
3.7	Erreur relative des réseaux RBF pour 3 types de fonction de coût	112
4.1	Résultats de l'identification des paramètres probabilistes de la courbe de Wöhler pour l'essieu X52 Ph1	119
4.2	Résultats de l'identification des paramètres probabilistes de la courbe de Wöhler pour l'essieu X52 Ph2	120
4.3	Résultats de l'identification des paramètres probabilistes de la courbe de Wöhler pour l'essieu X98	120
4.4	Coefficient de variation du paramètre a pour X52 Ph1	122
4.5	Coefficient de variation du paramètre a pour X52 Ph2	122
4.6	Coefficient de variation du paramètre a pour X98	122
4.7	Facteurs de corrélation pour l'essieu X52 Ph1	125
4.8	Facteurs de corrélation pour l'essieu X52 Ph2	125

Nomenclature

μ_a, σ_a	Paramètres de la loi Log-normale pour a / Parameters of Log-normal distribution for a
μ_t, σ_t	Paramètres de la loi Log-normale pour T / Parameters of Log-normal distribution for T
a	Paramètre de la loi de Basquin / Parameter of Basquin law
b	Paramètre pente de la loi de Basquin / Slope parameter of Basquin law
d	Domage / Damage
N^f	Nombre d'initiations de fissures / Number of crack initiations
n_i	Nombre de cycles d'un amplitude de contrainte / Cycle number for a stress amplitude
R_m	Résistance à la traction / Tensile strength
S_D	Limite d'endurance / Fatigue limit
S_i	Amplitude de contrainte / Stress amplitude
T	Durée de vie / Fatigue lifetime
α	Paramètre d'échelle de la loi de Weibull / Scale parameter of Weibull distribution
β	Paramètre de forme de la loi de Weibull / Shape parameter of Weibull distribution
μ	Paramètre de la loi Log-normale / Parameter of Log-normal distribution
μ_L	Moyenne de la loi Log-normale / Mean of Log-normal distribution
μ_W	Moyenne de la loi de Weibull / Mean of Weibull distribution
σ	Paramètre de la loi Log-normale / Parameter of Log-normal distribution
σ_L	Ecart-type de la loi Log-normale / Standard deviation of Log-normal distribution
σ_W	Ecart-type de la loi de Weibull / Standard deviation of Weibull distribution
c	Paramètre de position de la loi de Weibull / Position parameter of Weibull distribution
M_L	Médiane de la loi Log-normale / Median of Log-normal distribution
M_W	Médiane de la loi de Weibull / Median of Weibull distribution

$r.v.$	Variable aléatoire / Random variable
\log	Logarithme naturel / Natural logarithm
\mathbb{E}	Espérance / Expectation
\mathbb{P}	Probabilité / Probability
\mathcal{IG}	Loi inverse gamma / Inverse gamma distribution
\mathcal{LN}	Loi Log-normale / Log-normal distribution
\mathcal{L}	Fonction de vraisemblance / Likelihood function
\mathcal{NIG}	Loi normal-inverse-gamma / Normal-inverse gamma distribution
\mathcal{N}	Loi normale / Normal distribution
\mathcal{X}	Espace d'état de chaîne de Markov / State space of Markov chain
π_{pr}	Loi <i>a priori</i> pour estimation Bayésienne / Prior distribution for Bayesian estimation
$\xrightarrow{a.s.}$	Convergence presque sûre / Almost sure convergence
\xrightarrow{D}	Convergence en loi / Convergence by distribution
F	Fonction de répartition / Cumulative distribution function
f	Densité de probabilité / Probability density function
R	Fiabilité / Reliability
IS	Importance sampling
MCMC	Monte Carlo Markov chain
MH	Metropolis-Hastings
MLE	Maximum likelihood estimation
PMF	Fonction de masse / Probability mass function
ω	Poids des neurones en couche linéaire / Weights in the linear output neuron
\mathbf{c}	Centre de RBF / RBF center
ω_n^2	Statistique du test de Cramér-von Mises / Statistic of Cramér-von Mises test
ω_{n_1, n_2}	Statistique du test à deux échantillons de Cramér-von Mises / Statistic of two-sample Cramér-von Mises test
D^2	Statistique du test de χ^2 / Statistic of χ^2 test
D_n	Statistique du test de Komlogorov-Smirnov / Statistic of Komlogorov-Smirnov test
D_{n_1, n_2}	Statistique du test à deux échantillons de Komlogorov-Smirnov / Statistic of two-sample Komlogorov-Smirnov test
e_{μ_a}	Erreur relative de μ_a identifié par rapport à la valeur de référence / Relative error of identified μ_a identifié with respect to the reference value

F_n^* Fonction de répartition empirique / Empirical distribution function

CF Fonction de coût / Cost function

RBF Réseau à fonctions à base radiale / Radial basis function

Introduction générale

Les dommages dûs à la fatigue mécanique sont l'une des sources les plus importantes de défaillance des structures [1], notamment dans les industries du transport, aéronautique, ferroviaire et automobile. Dans l'industrie automobile, la prise en compte de la fiabilité en fatigue lors de la conception des structures est souvent basée sur des approches combinant à la fois des essais expérimentaux et des simulations numériques. La validation de la conception en terme de fatigue est effectuée pour différents niveaux de structures : la voiture complète, les sous-systèmes et les organes. La validation progressive, de l'organe à la voiture complète, permet de réduire le temps et le coût de conception en détectant les défaillances avant l'intégration d'un élément dans un système plus important.

Dans le contexte industriel du constructeur automobile Renault, l'essai expérimental à l'échelle d'une voiture complète durant tout son cycle de vie n'est pas possible à cause de sa longue durée. Afin de réduire cette durée, des essais accélérés ont été développés. Ils utilisent des chargements plus sévères que la demande du client (profil de mission) et permettent donc d'obtenir des défaillances prématurées. Ces essais accélérés sont conduits pour les différents niveaux de structures dans le cadre de la validation de la conception. Une part toujours plus importante est accordée aux simulations numériques, représentatives des essais physiques, afin de pouvoir remplacer les essais graduellement. Les simulations numériques permettent en effet de réduire le coût de la validation, notamment en terme de temps.

La fiabilité en fatigue est un phénomène aléatoire difficile à déterminer pour les structures mécaniques, surtout pour les structures complexes automobiles. Les modes de défaillances sont multiples, il existe de nombreux facteurs d'influence : comportement du matériau, géométrie, assemblage, procédés de fabrication et incertitudes du chargement [2]. En termes de défaillances, les données expérimentales sont peu nombreuses car les essais physiques sont coûteux. De plus, certaines données sont souvent censurées, l'information est moins riche lorsque l'instant de défaillance n'est pas défini. Ce manque d'information est donc un frein à une bonne estimation de la fiabilité. Une partie du travail de thèse portant sur l'estimation des données expérimentales a pris en compte ces données censurées dans le modèle de fiabilité.

Dans le cadre de la validation des châssis Renault, les développements proposés dans cette thèse visent à prédire la variabilité de la durée vie, le modèle doit donc être non déterministe. L'objectif principal de ce travail de thèse est d'introduire la variabilité dans la simulation numérique grâce à un modèle probabiliste. Ce travail doit permettre à terme de mieux interpréter les résultats d'essais et de réduire leur nombre si le modèle probabiliste est intégré dans le processus de validation.

Les travaux de thèse se focalisent sur les essais et la simulation numérique des organes du châssis automobile. La variabilité est introduite dans la modélisation numérique en supposant qu'elle est principalement liée à la courbe de Wöhler, soit encore aux résultats d'essais de fatigue. L'identification d'un modèle probabiliste de la courbe de Wöhler est donc visée. Pour cela, une méthode d'identification est proposée, basée sur une approche inverse utilisant les résultats d'estimation des données expérimentales.

En considérant le contexte de la validation des châssis Renault et les objectifs précédents visant à l'améliorer, le manuscrit est organisé de la façon suivante :

- Chapitre 1 : Le contexte du projet est présenté, comprenant le processus de validation, les pièces étudiées, un état de l'art sur les essais physiques et les simulations numériques chez Renault. Les notions de fiabilité en fatigue, le dommage et la durée de vie sont introduites. La courbe de Wöhler, propriété importante du matériau en fatigue, ainsi que des modèles déterministe et probabiliste sont présentés.
- Chapitre 2 : Plusieurs méthodes et modèles d'estimation des données censurées permettant de quantifier la variabilité des durées de vie expérimentales sont proposés. Dans un premier temps, la notion de données censurées de type groupées est présentée. La méthode d'estimation du maximum de vraisemblance est appliquée sur les données groupées avec des modèles basés sur les lois Log-normale et de Weibull. Le calcul de l'intervalle de confiance avec des données groupées est démontré. Ensuite, l'estimation Bayésienne est proposée pour prendre en compte l'information *a priori* des paramètres pour la loi Log-normale et la loi de Weibull. Les lois *a priori* sont choisies comme des lois conjuguées ou non informatives. En plus des lois continues Log-normale et de Weibull, une loi discrète multinomiale est proposée pour décrire les données groupées. Enfin, une série de données simulées avec une loi de référence est utilisée pour quantifier les performances des différentes méthodes d'estimation et des différents modèles.
- Chapitre 3 : Dans ce chapitre, une stratégie d'identification des incertitudes liées à la courbe de Wöhler est proposée avec une approche analytique et une approche par la simulation. La stratégie de l'approche par la simulation est construite en deux étapes. La première étape correspond à l'estimation de la durée de vie expérimentale développée dans le chapitre 2. La deuxième étape vise à minimiser la distance entre la variabilité de la durée de vie expérimentale et la variabilité de la durée de vie obtenue à l'issue de la simulation éléments finis. La durée de vie dite éléments finis est générée avec une courbe de Wöhler probabiliste, dont les paramètres sont à identifier. La distance entre les variabilités est mesurée à l'aide de fonctions de coût proposées. Une étude de convergence en fonction de la taille de l'échantillon de la durée de vie éléments finis est menée sur les fonctions coût. Également, une étude sur la précision d'identification du modèle liée à la censure des données groupées et sur l'erreur liée à la méthode d'estimation est réalisée à l'aide d'un réseau à fonctions de base radiale.
- Chapitre 4 : L'application des travaux de thèse au contexte industriel est proposée. La stratégie d'identification de la variabilité liée à la courbe de Wöhler est appliquée sur les données réelles des projets Renault X52 et X98. La courbe de

Wöhler probabiliste est identifiée pour chaque cordon de soudure des deux projets. Les résultats obtenus avec les différentes méthodes sont comparés. Avec la courbe de Wöhler probabiliste identifiée, un calcul probabiliste de la durée de vie est proposé. De plus, les résultats permettent de donner un nouveau critère de validation du calcul éléments finis. Pour faciliter l'application de la méthodologie, une interface sous MATLAB a été développée, elle est présentée en annexe.

- Une conclusion générale synthétise les résultats présentés dans ce rapport de thèse dans le contexte de la validation des châssis automobiles. Des perspectives à ces travaux sont également proposées.

Les travaux de thèse ont été réalisés dans le cadre d'une collaboration entre le constructeur automobile Renault et l'Université de Technologie de Compiègne (UTC), avec un financement de type CIFRE. Pour Renault, les équipes déploiement et simulation numérique ont été impliquées. Pour l'UTC, la co-direction de thèse a été partagée entre le laboratoire de mécanique Roberval et le laboratoire de mathématiques appliquées LMAC.

Chapitre 1

État de l'art sur la fatigue des structures automobiles

Dans ce premier chapitre du manuscrit, le contexte industriel automobile de la thèse est tout d'abord présenté. Il est composé du processus de validation des châssis, d'une présentation des structures étudiées, d'une description des essais physiques et des simulations numériques sur les châssis Renault. Un état de l'art de la fatigue mécanique est ensuite proposé, les notions de dommage et de durée de vie sont introduites, la représentation des essais de fatigue sous la forme de la courbe de Wöhler est décrite, ainsi que des modèles déterministes et probabilistes associés à ce comportement.

Contents

1.1 Validation en fiabilité et tests d'endurance	6
1.1.1 Objectif de fiabilité	6
1.1.2 Cycle en V de la validation	7
1.1.3 Trois essais d'endurance	8
1.1.3.1 Essai de roulage	8
1.1.3.2 Essai dynamique	9
1.1.3.3 Essai synthétique	9
1.1.4 Liaison au sol	10
1.1.5 Essai synthétique sur essieu souple et mesure de durée de vie .	12
1.1.6 Simulation numérique de l'essai synthétique	13
1.2 Modélisation de la fatigue	14
1.2.1 Calcul du dommage sous chargement avec amplitude variable	14
1.2.2 Modèle déterministe de la courbe de Wöhler	18
1.2.3 Modèle probabiliste de la courbe de Wöhler	19
1.3 Conclusion	20

1.1 Validation en fiabilité et tests d'endurance

Dans l'industrie automobile, avant qu'un véhicule soit lancé en production sur les chaînes de montage, la validation de la conception demande plusieurs itérations de modifications. La validation en terme de fatigue a pour objectif la fiabilité. Cette validation est effectuée à différentes échelles : la voiture complète, les sous-systèmes et les composants. Pour ces trois types de structures, des tests d'endurance et des simulations numériques équivalentes sont menées pour contribuer à la validation.

1.1.1 Objectif de fiabilité

L'étude de fiabilité est une étude des défaillances des systèmes et des composants d'un point de vue statistique. La fiabilité au temps t est la probabilité qu'un composant, un système ou un sous-système ne subisse pas de panne dans l'intervalle de temps $[0, t]$. D'autres indicateurs de fiabilité sont le temps moyen de bon fonctionnement du produit avant la première défaillance, le taux de défaillance, et le temps moyen entre deux défaillances. Pour un véhicule, le client s'intéresse souvent à la durée de fonctionnement en fonction du nombre de kilomètres qu'il va pouvoir parcourir avant défaillance. Le kilométrage tient ici lieu de temps. La performance d'endurance d'un produit est liée à la fiabilité en fatigue. La conception des véhicules intègre également la résistance au choc, le confort et le comportement ; plusieurs tests de validation sont dédiés aux aspects crash. Pour certains composants automobiles, la fatigue causée par le chargement cyclique thermique ou l'humidité constituent également des tests de validation importants. Dans le cadre de ce mémoire de thèse, notre attention se situe uniquement dans le périmètre de l'endurance mécanique.

La fiabilité en endurance nécessite une approche probabiliste, car la fatigue des matériaux est de nature probabiliste. Le dommage dû à la fatigue est l'une des sources les plus importantes de défaillance des structures mécaniques [1], notamment dans l'industrie automobile. Zoroufi et Fatemi [3] introduisent la prédiction de la durée de vie d'un porte-fusée de châssis à l'aide d'essai expérimental et de modèles numériques. Sonsino et al. [4] proposent une étude sur la résistance à la fatigue des arbres d'entraînement des boîtes de vitesses automobiles pour des couples de fonctionnement. La fatigue est un processus d'endommagement local d'un composant ou d'un système soumis à un chargement cyclique, comprenant l'initiation et la propagation de fissure jusqu'à la rupture [5]. D'un point de vue microscopique, la structure d'un matériau est rarement homogène [6, 7] ; des défauts situés aléatoirement peuvent initier des micro fissures. La propagation de fissure est également un aspect probabiliste que de nombreux travaux ont tenté de modéliser [8, 9]. Schijve et al. [2] ont également étudié les effets de taille de la structure, l'influence des paramètres de production et les spectres des chargements. Étant données les nombreuses sources d'incertitude dans l'étude d'endurance, une approche probabiliste est indispensable. Les critères de validation en fatigue des châssis automobiles doivent donc être satisfaits avec une approche probabiliste.

Pour satisfaire la demande client (profil de mission) en terme de fiabilité, le critère de validation est souvent plus sévère que cette demande, ce qui induit un coefficient de

sécurité. Pour les trois niveaux de structure, les coefficients de sécurité sont différents. Le coefficient de sécurité pour un composant est souvent plus élevé que pour un système mécanique, car les composants ont une relation en série dans un système, qui conduit à une fiabilité pour un système souvent plus faible que pour un composant.

La méthode contrainte-résistance est une approche statistique qui permet de relier la capacité à résister aux chargements (résistance) avec les sollicitations internes (contraintes) d'un système auxquelles il est soumis. Si on considère que l'endommagement lié à la fatigue du profil de mission correspond aux «contraintes», et que l'endommagement demandé par un test de validation correspond à la «résistance», on peut décrire cette relation avec la Figure 1.1. Dans une approche déterministe, des indicateurs peuvent être utilisés pour définir le coefficient de sécurité, par exemple les médianes des densités de probabilité de la fiabilité. Des approches probabilistes peuvent être appliquées pour l'analyse de fiabilité, comme les méthodes FOSM (First Order Second Moment) et FORM (First Order Reliability Method). Les objectifs des tests de validation sont alors atteints, et servent à établir un cahier des charges (CDC) des tests de validation. Dans le vocabulaire employé chez Renault, le CDC définit les sollicitations d'un essai. Dans un contexte similaire, Thomas et al. [10] ont mené un travail pour définir un objectif de fiabilité des composants en utilisant la méthode contrainte-résistance dans l'industrie automobile. Laribi et al. [11] ont également proposé une nouvelle approche probabiliste de fiabilité pour appliquer de façon appropriée la méthode de contrainte-résistance.

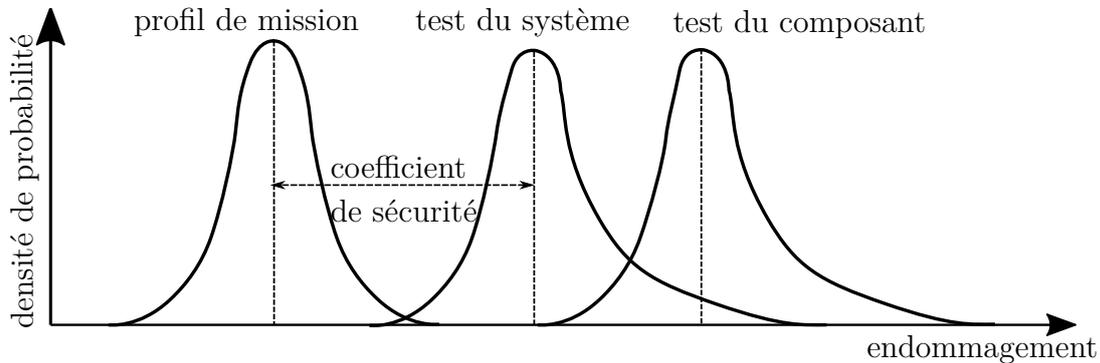


FIGURE 1.1 – Relation entre l'objectif de validation du profil de mission et les tests sur système ou composant

1.1.2 Cycle en V de la validation

La conception dans le cadre de l'endurance nécessite des choix appropriés de matériaux, de géométries et de dimensions pour que la structure permette d'atteindre l'objectif en fiabilité définie à partir du profil de mission. Le CDC représente un objectif en fiabilité, la structure doit donc pouvoir résister aux chargements définis par ce CDC afin que sa résistance en fatigue soit valide. La validation d'une structure est satisfaite en réalisant les essais expérimentaux définis par le CDC. Les indicateurs de ces essais sont la durée de vie de la structure (composant, système), et également les zones critiques où les fissures apparaissent. Les procédures de validation de conception en terme d'endurance chez Renault peuvent être décrites par un cycle en V présenté par la Figure 1.2.

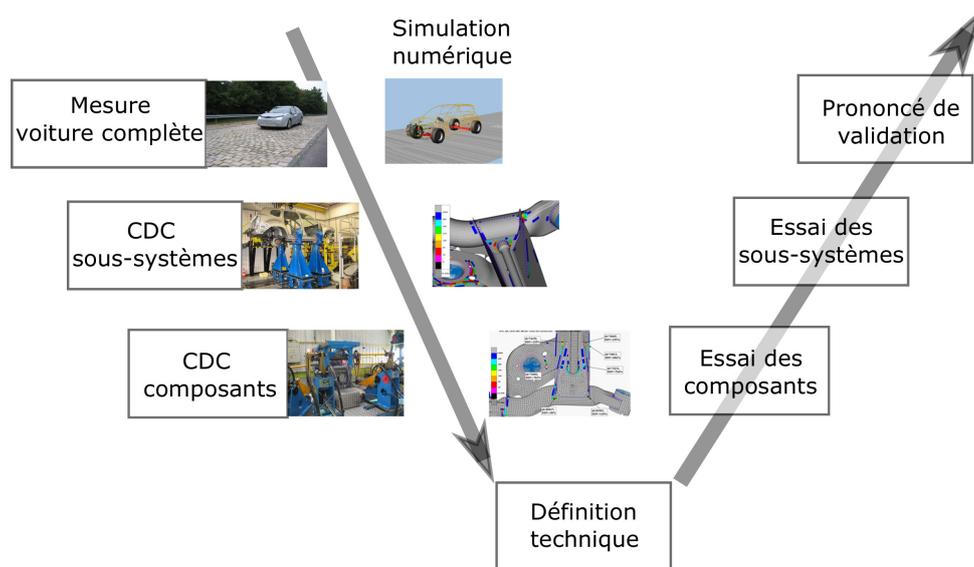


FIGURE 1.2 – Cycle en V pour la validation du déploiement

Lors d'un nouveau projet, le CDC de l'ancien projet est repris, les trois niveaux d'essai sont réalisés, ainsi que les simulations numériques correspondantes. Il faut noter que la validation est principalement faite par essais physiques, elle peut s'accompagner de simulation numérique. Puis, la définition technique des pièces ou systèmes est validée par les essais. Avant la fabrication en série, les essais dynamiques sur sous-systèmes et les essais synthétiques sur composants sont relancés, des petites modifications sont faites si besoin, jusqu'à prononcer la validation finale. Le CDC doit s'adapter au cours du processus de validation, il est difficile de définir les chargements pour atteindre un objectif de fiabilité sans réaliser d'essai. A partir du CDC d'un ancien projet, il est amélioré avec le retour du banc d'essai pour s'adapter au nouveau projet. La validation du châssis est un travail de coopération entre différentes équipes de plusieurs sites : les équipes essais, l'équipe déploiement déterminant le CDC pour chaque essai et l'équipe calcul réalisant les modélisations et simulations. Dans un contexte similaire, un travail de thèse a mené à l'amélioration du protocole des essais d'endurance [12].

1.1.3 Trois essais d'endurance

Trois types d'essais d'endurance et de fatigue sont réalisés sur différents niveaux de structure : essai de roulage pour une voiture complète, essai dynamique pour les sous-systèmes et essai synthétique pour les composants. Ces essais visent à valider la durabilité et la fiabilité des structures et des composants automobiles.

1.1.3.1 Essai de roulage

L'essai de roulage présenté dans la Figure 1.3 est réalisé sur la piste de référence du site de Lardy. Cette piste comprend les sections des différents profils de route reproduisant des sollicitations plus sévères qu'une route classique (profil de mission) pour assurer le dommage sur une distance limitée. Un véhicule expérimental équipé de capteurs roule



FIGURE 1.3 – Essai de roulage

sur la piste, et les signaux (efforts et moments sur roues, débattement vertical) mesurés alimentent la définition du CDC des essais sur sous-systèmes et composants.

1.1.3.2 Essai dynamique

L'essai dynamique (Figure 1.4) est représentatif de l'essai de roulage sur piste. Il tente de reproduire le plus fidèlement possible les signaux des sollicitations mesurées lors du roulage, avec 6 degrés de liberté (3 efforts et 3 moments) sur chaque roue. Cet essai de synthèse a pour but de valider en endurance les sous-systèmes, comme une caisse en blanc, un train avant ou un train arrière.



FIGURE 1.4 – Essai dynamique

1.1.3.3 Essai synthétique

L'essai synthétique est un essai quasi-statique sur un composant qui a pour objectif principal de reproduire de manière accélérée les défaillances observées sur l'essai dynamique. La durée de cet essai est plus courte que celle de l'essai dynamique. De plus comme l'essai est quasi-statique, la fréquence des signaux est basse, seules les sollicitations les plus endommageantes sont conservées. Avant l'essai, la structure instrumentée par jauges et

rosettes présentée dans la Figure 1.5a est chargée pour relever l'état de déformation près des cordons de soudure et des bords de découpes. Les déformations mesurées servent à définir le CDC pour atteindre un objectif de fiabilité. L'essai synthétique est effectué sous sollicitations cycliques jusqu'au dysfonctionnement du composant.



(a) Essieu instrumenté avec rosettes



(b) Essai synthétique

FIGURE 1.5 – Essieu instrumenté avant l'essai et essai synthétique

Le banc de l'essai synthétique pour un essieu comporte 6 vérins (Figure 1.5b), 3 degrés de liberté sur chaque roue, permettant de piloter les efforts cycliques longitudinal, latéral et le déplacement vertical des deux côtés de l'essieu. Afin de déterminer la variabilité de la durée de vie en fatigue, l'essai synthétique est réalisé pour différentes structures théoriquement identiques de même conception. L'essieu arrière est relié sur le banc d'essai avec une articulation, des amortisseurs et des ressorts.

En terme de fiabilité d'endurance, ces trois types d'essai sont utilisés dans le processus de validation. D'autres essais complémentaires de choc et de crash sont également réalisés pour la validation de la conception des composants et des systèmes, mais ne sont pas traités dans cette thèse.

1.1.4 Liaison au sol

La liaison au sol désigne l'ensemble des composants qui effectuent le lien entre le châssis d'un véhicule et la route. Elle supporte le véhicule et assure la rotation des roues. Cette liaison assure la transmission des efforts moteur et de freinage, le guidage et le confort dans des conditions de sécurités maximales. Elle est constituée par le système de direction, les suspensions, les roues et les trains roulants. La liaison au sol du modèle Clio 4 2019 est présentée dans la Figure 1.6.

Les trains roulants avant et arrière (Figure 1.7) permettent de relier la roue au châssis, d'assurer le guidage et le freinage du véhicule. Le composant d'intérêt de la thèse est le train arrière, qui est souvent un essieu souple.

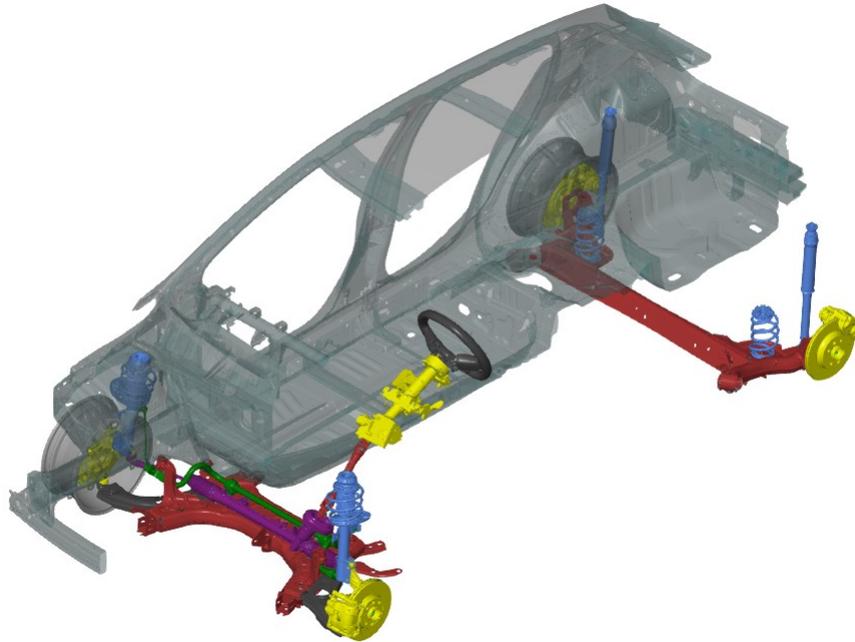
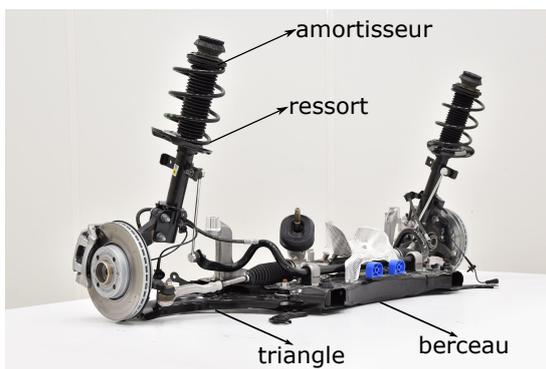
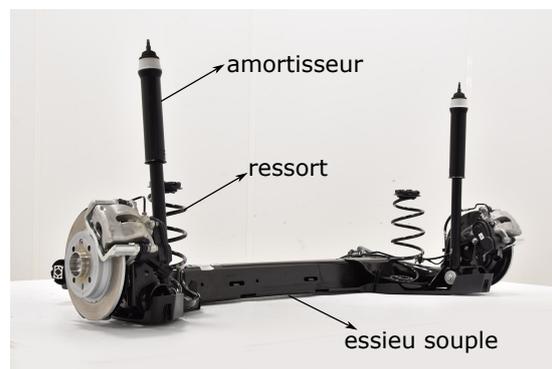


FIGURE 1.6 – Système liaison au sol (Clio 4 2019)



(a) Train avant



(b) Train arrière

FIGURE 1.7 – Trains roulants

1.1.5 Essai synthétique sur essieu souple et mesure de durée de vie

L'essai synthétique est appliqué aux composants, tels que essieu souple, berceau, triangle, porte fusée, etc. Nous nous focalisons sur l'essai synthétique d'essieu souple pour ces travaux de thèse. Le banc d'essai est présenté sur la Figure 1.5b et la structure testée sur la Figure 1.8.

L'essai vise à obtenir les données expérimentales composées de la durée de vie en fatigue pour chaque joint de soudure du composant. Cette durée de vie expérimentale en fatigue d'un joint de soudure est définie entre le début de l'essai jusqu'à l'apparition de microfissures observées. Ces fissures sont détectées avec un contrôle par ressuage (Figure 1.9). L'inspection nécessite d'interrompre plusieurs fois l'essai, elle n'est donc pas effectuée de manière continue dans le temps, mais de manière discrète. À chaque inspection, la longueur de fissure observée est mesurée et le temps d'interruption enregistré.

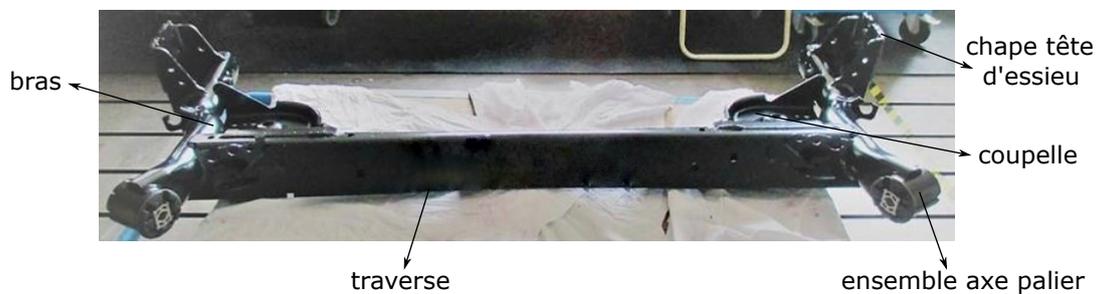


FIGURE 1.8 – Essieu souple automobile

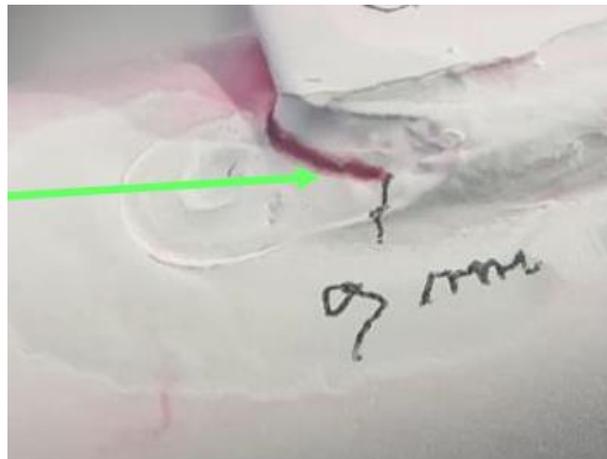


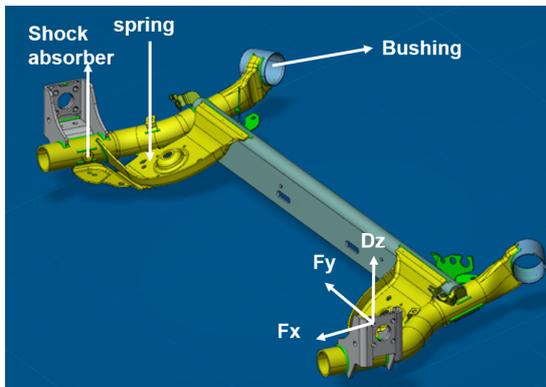
FIGURE 1.9 – Fissure proche d'un cordon de soudure identifiée par ressuage

L'essai synthétique est réalisé sur des pièces théoriquement identiques. L'analyse des résultats expérimentaux permet de quantifier une variabilité due aux procédés de fabrication et aux matériaux. Ainsi, la durée de vie en fatigue doit être considérée comme une variable aléatoire déterminée par une description probabiliste. Cette description peut être réalisée grâce à la distribution probabiliste des données qui est généralement estimée par des méthodes paramétriques ou non-paramétriques. Il a été mentionné que l'inspection

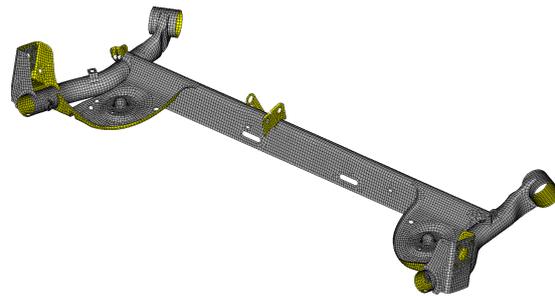
par ressuage demande des interruptions durant l'essai. Les fissures sont détectées lors d'une inspection, elles apparaissent donc entre deux instants de ressuage. Par conséquent, le moment exact de fissurations n'est pas observé, en revanche le nombre de défaillances dans des intervalles d'inspection est obtenu. Ces interruptions conduisent à des données groupées pour la durée de vie expérimentale en fatigue. Cela signifie que le temps d'initiation d'une fissure n'est pas observé, mais que le nombre de fissures est compté dans chaque intervalle de temps d'interruption. Les données groupées nécessitent donc un traitement adapté afin que la description probabiliste puisse être établie. Dans les développements de thèse proposés, une méthode d'estimation paramétrique pour les données groupées est proposée et présentée dans le chapitre 2. La valeur médiane B50 de la durée de vie, souvent utilisée dans l'industrie, est obtenue par l'estimation.

1.1.6 Simulation numérique de l'essai synthétique

Le modèle éléments finis de la structure étudiée est préparé et fourni par le Groupe Renault. L'essieu arrière du châssis est un assemblage de pièces de tôlerie, principalement modélisé par des éléments finis de coque. Les cordons de soudure sont modélisés à l'aide d'une technique spécifique. Les composantes du chargement (F_x , F_y) du banc d'essai et le déplacement vertical (D_z) sont indiqués sur la Figure 1.10(a). Les efforts de réaction du ressort et de l'amortisseur sont également pris en compte. Des conditions aux limites sont appliquées pour modéliser les articulations. Le cycle du chargement est basé sur celui de l'essai expérimental. La résolution éléments finis est effectuée avec le solveur Nastran dans le cadre de l'analyse statique.



(a) Modèle CAO



(b) Maillage d'un essieu

FIGURE 1.10 – Modèle CAO et maillage éléments finis d'un essieu

Les contraintes obtenues de l'analyse par éléments finis sont utilisées pour calculer le dommage lié à la fatigue et la durée de vie de chaque joint de soudure. Le dommage est évalué sur un élément fini approprié proche du cordon de soudure. Le calcul du dommage et de la durée de vie est présenté dans la section suivante. Le dommage est évalué ici numériquement dans un cadre déterministe, qui ne tient donc pas compte des incertitudes observées en terme de la durée de vie expérimentale.

1.2 Modélisation de la fatigue

Les contraintes calculées au voisinage des cordons de soudure alimentent le calcul du dommage et de la durée de vie. Pour un assemblage soudé, les joints de soudure sont des zones critiques où les premières fissures peuvent apparaître [13]. Le processus de soudage conduit en effet à une structure du matériau non homogène et à des défauts, ce qui induit des concentrations de contraintes et des contraintes résiduelles [14]. En pratique, les essais expérimentaux montrent toujours un niveau élevé de variabilité en terme de durée de vie en fatigue. La connaissance du comportement en fatigue des joints de soudure est donc importante pour la prédiction de la durée de vie de la structure.

Cependant, la simulation numérique menant à la durée de vie actuellement pratiquée par le partenaire industriel est déterministe. Nous souhaitons donc proposer une méthodologie pour une simulation probabiliste de la fatigue. En calcul de structure, les sources d'incertitudes sont diverses : chargement, propriétés matériaux, données expérimentales, ou encore incertitudes concernant les choix de modélisation numérique [15]. Nous avons choisi de formaliser via la courbe de Wöhler un ensemble de dispersions dans les travaux de la thèse. Le calcul du dommage cumulé, ainsi que les modèles déterministe et probabiliste de la courbe de Wöhler sont présentés dans les sections suivantes.

1.2.1 Calcul du dommage sous chargement avec amplitude variable

Pour caractériser le comportement d'un matériau en fatigue, les éprouvettes sont généralement soumises à des chargements avec amplitude constante, pour obtenir notamment la courbe de Wöhler. Mais les essais de validation sur châssis ont des chargements avec des amplitudes variables pour être représentatifs de conditions de roulage. En fait, l'ordre des cycles de contrainte du chargement peut influencer le comportement du matériau. Zakaria et al. [16] présentent différentes images obtenues par fractographie de stries de fatigue pour des chargements avec amplitudes constante et variable. Les auteurs indiquent également que l'effet d'une amplitude variable est essentiel pour une approche basée sur la déformation (fatigue à nombre de cycles faible) conduisant à la déformation plastique. Mais avec une approche basée sur la contrainte (fatigue à grand nombre de cycles), l'effet d'une amplitude variable est moins évident. Dans notre étude, pour évaluer le dommage et la durée de vie sous chargement avec amplitude variable, les hypothèses suivantes sont faites [17] :

- l'ordre des cycles de contrainte n'influence pas le cumul du dommage,
- la fréquence du chargement n'influence pas la fatigue,
- le dommage causé par un cycle peut être traité individuellement.

Le dommage lié à la fatigue est défini par le rapport entre le nombre de cycles n_i pour une amplitude de chargement S_i et son nombre critique de cycles jusqu'à la rupture N_i déterminé par la courbe de Wöhler. Pour obtenir ce comportement, des éprouvettes homogènes sont sollicitées avec un chargement périodique caractérisé par sa contrainte moyenne et sa contrainte alternée, le nombre de cycles à la rupture du matériau est enregistré. La courbe décrivant la relation entre le nombre de cycles à rupture et une

amplitude de la contrainte alternée se nomme courbe de fatigue, courbe de Wöhler ou encore courbe S-N. Cette courbe est présentée sur la Figure 1.11, avec R_m la résistance à la traction et S_D la limite d'endurance. La représentation des résultats d'essais de fatigue met en évidence classiquement deux régions sur la courbe : une région oligocyclique pour la fatigue à faible nombre de cycles, et une région conventionnelle pour la fatigue à grand nombre de cycles. La partie conventionnelle peut être séparée en deux domaines, l'endurance limitée et l'endurance illimitée [18]. La limite d'endurance S_D est mise en évidence lorsque l'amplitude de la contrainte est inférieure à la limite de fatigue. Avec cette amplitude de chargement (contrainte), la durée de vie en fatigue est supposée être infinie.

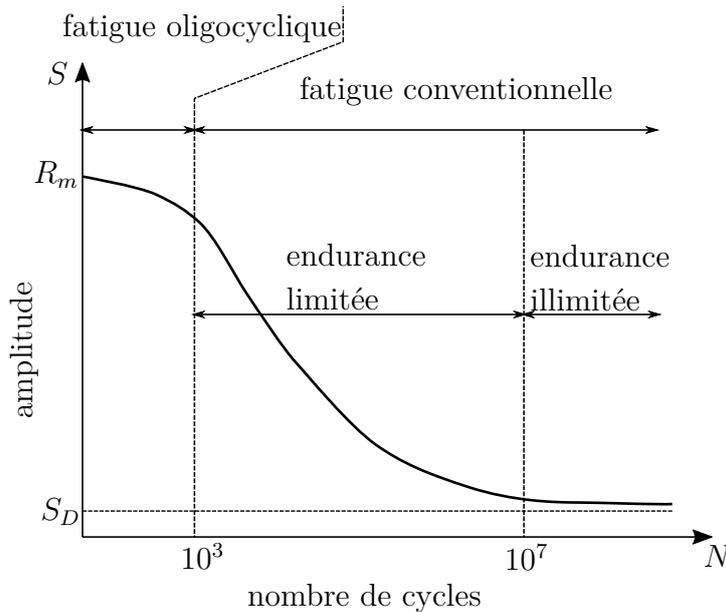


FIGURE 1.11 – Courbe de Wöhler

Dans la zone oligocyclique, la déformation est souvent utilisée au lieu de la contrainte (modèle de Manson-Coffin), car la déformation plastique apparaît dans cette région. La loi de comportement du matériau est non-linéaire, avec boucles d'hystérésis, le phénomène peut être modélisé par une méthode de dissipation d'énergie. La région d'intérêt dans notre étude est la région d'endurance limitée pour la fatigue à grand nombre de cycles, car les chargements opérationnels durant le roulage de la voiture n'induisent pas de déformation plastique en condition normale. Cette région correspond souvent à un nombre de cycles compris entre 10^3 et 10^6 , ou 10^3 et 10^7 selon les matériaux. Dans notre étude, un nombre de cycles égal à 10^7 en borne supérieure de la région est considéré.

À partir des contraintes temporelles obtenues sur un élément fini proche d'un joint de soudure, la méthode de comptage Rainflow [19] est utilisée pour obtenir les amplitudes et valeurs moyennes des contraintes et le nombre de cycles correspondant. Son efficacité a été démontrée pour le comptage des cycles dans l'évaluation des dommages. L'algorithme standard de la méthode Rainflow est décrit par Amzallag et al. [20] et la norme ASTM [21].

Il faut noter que si la moyenne du chargement cyclique est différente de 0, les méthodes de correction, par exemple la méthode de Goodman ou Gerber, peuvent être uti-

lisées pour corriger l'effet de la moyenne [22]. Quand la moyenne du chargement n'est pas nulle, l'effet est différent. On considère que l'effet du chargement avec son amplitude de contrainte alternée S et la moyenne du chargement σ_m est la somme de l'effet de S et σ_m (Figure 1.12). Une valeur moyenne positive induit un préchargement en traction, avec un effet négatif sur le comportement en fatigue réduisant le nombre de cycles avant rupture. Une valeur moyenne négative induit de la compression, avec un effet positif sur le comportement en fatigue. Ces effets sont présentés sur la courbe de Wöhler de la Figure 1.13.

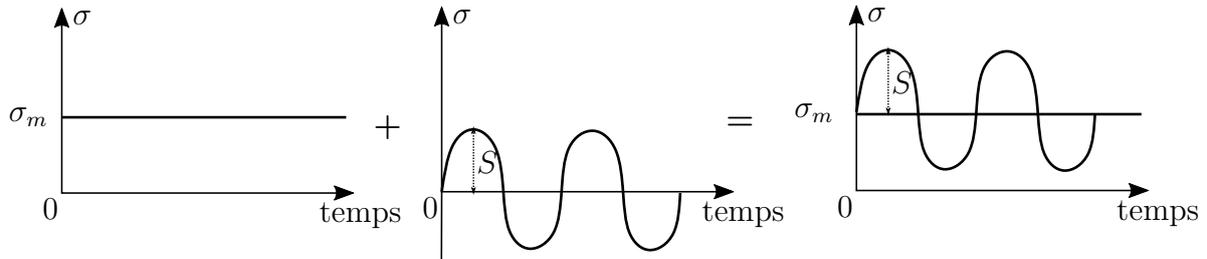


FIGURE 1.12 – Superposition de la moyenne et amplitude de contrainte

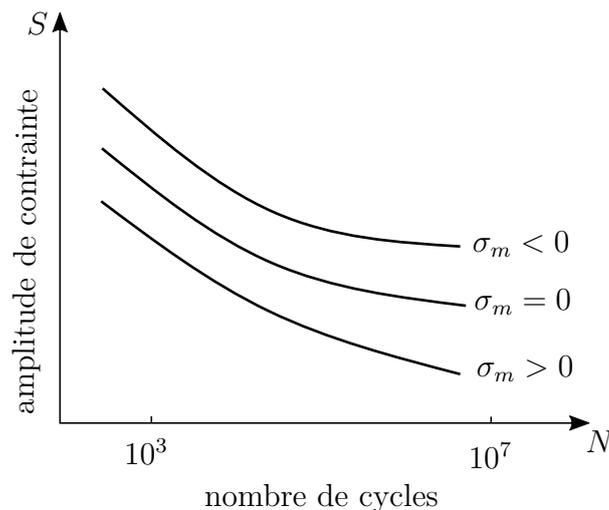


FIGURE 1.13 – Influence de la valeur de contrainte moyenne du chargement cyclique sur la courbe de Wöhler

L'influence d'une contrainte moyenne non nulle devra donc être corrigée dans le calcul de fatigue. Pour cela, l'idée est de trouver l'amplitude de la contrainte purement alternée qui causera un dommage équivalent au chargement à contrainte moyenne nulle. Plusieurs méthodes ont été proposées pour donner la correspondance entre la contrainte moyenne σ_m et l'amplitude alternée S . Ici deux méthodes sont introduites, car elles sont couramment utilisées dans l'industrie. Les diagrammes de Goodman et Gerber sont présentés sur la Figure 1.14. La correction de Goodman propose une droite avec une coordonnée à l'origine correspondant à la limite d'endurance S_D (ou encore la résistance à la fatigue pour un nombre de cycle donné), l'intersection avec l'axe des abscisses correspond à la résistance à la traction, cette fonction est décrite par l'équation suivante,

$$\frac{S}{S_D} + \frac{\sigma_m}{R_m} = 1. \quad (1.1)$$

Le diagramme de Goodman reste valide dans la partie compression avec $\sigma_m < 0$. Le diagramme de Gerber est quant à lui une représentation parabolique passant par les points précédemment décrits. Sa fonction est décrite par l'équation,

$$\frac{S}{S_D} + \left(\frac{\sigma_m}{R_m}\right)^2 = 1, \sigma_m > 0. \quad (1.2)$$

La représentation de Gerber ne peut pas s'étendre à la région de compression, avec un effet négatif en fatigue qui n'a pas de sens. En pratique, l'effet de compression n'est généralement pas considéré.

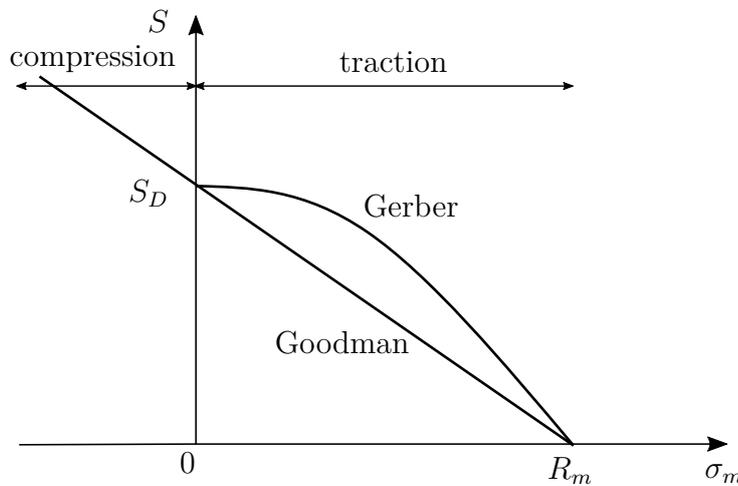


FIGURE 1.14 – Corrections de Goodman et Gerber

Le domaine situé au dessous des diagrammes de Goodman et Gerber est une région de sécurité. Pour une moyenne de contrainte et une amplitude de contrainte correspondant dans cette région, la défaillance liée à la fatigue ne se produira pas. De ce point de vue, le diagramme de Goodman est plus conservatif que celui de Gerber, en terme de critère de conception en considérant l'effet de la moyenne de contrainte [23]. Les données expérimentales des matériaux ductiles sont souvent associées à une description de type Gerber, tandis que les tests des éprouvettes entaillées sont plus proches de la relation de Goodman. Étant donnée la dispersion observée en fatigue, il est préférable de faire un choix conservatif de conception avec la relation de Goodman. En pratique, le diagramme de Goodman est souvent utilisé dans l'industrie aéronautique et celui de Gerber dans l'industrie automobile.

Concernant le dommage cumulé, il est calculé en utilisant la règle d'accumulation linéaire de Miner [5] présentée par l'équation,

$$d = \sum_{i=1}^r \frac{n_i}{N_i}. \quad (1.3)$$

Avec d le dommage, n_i le nombre de cycles opérés, N_i le nombre maximum de cycles à rupture pour cette amplitude de contraintes, r le nombre d'amplitudes de contrainte

différentes comptées par l'approche Rainflow. Le dommage s'exprime avec la notion de fraction, pour la i^{me} amplitude de contrainte, le dommage $d_i = \frac{n_i}{N_i}$ représente la fraction de durée de vie consommée. Quand la fraction de durée de vie d_i cumulée est égale à 1, la durée de vie est consommée. En général, on considère que la défaillance se produit lorsque le dommage atteint la valeur 1. Ainsi, la durée jusqu'au dommage complet est la durée de vie définie par l'équation (1.4). Elle peut être évaluée numériquement suite à l'analyse éléments finis. Dans le cadre de ce travail de thèse, l'unité de la durée de vie est exprimée en pourcentage par rapport à la durée du CDC.

$$T = \frac{1}{d}. \quad (1.4)$$

D'autres règles d'accumulation sont présentées par Zhao et al. [24], mais il s'avère que l'hypothèse d'accumulation linéaire des dommages de Miner est adaptée à l'estimation de la durée de vie sous des charges d'amplitude variable aléatoire.

Le calcul du dommage précédent, utilisant la courbe de Wöhler, n'est normalement valable que pour des cas de chargement uni-axial. Pour des essais sur structure avec un chargement dans les trois directions de l'espace, le composant subit un état de contrainte multiaxiale. Plusieurs critères de fatigue multiaxiale ont été développés [25], comme par exemple le critère de Dang Van, mais il est souvent difficile à appliquer dans un contexte industriel. L'approche choisie ici est celle du plan critique [26]. Dans le modèle d'essieu, les éléments finis utilisés sont principalement de type coque sous hypothèse de contrainte plane. Pour un élément fini, le dommage induit par une série de chargements est calculé avec les contraintes dans plusieurs directions. La direction la plus endommagée est choisie, correspondant à un angle critique.

Basé sur les équations précédentes, le dommage et la durée de vie en fatigue sont calculés pour chaque joint de soudure selon le plan critique et la loi d'accumulation linéaire de Miner. La courbe de Wöhler joue un rôle essentiel dans la détermination du dommage et de la durée de vie des joints de soudure, évalués ici suite à une analyse éléments finis.

1.2.2 Modèle déterministe de la courbe de Wöhler

La courbe de Wöhler peut être représentée sous une forme linéaire sur une échelle logarithmique, comme illustré sur la Figure 1.15. Plusieurs modèles peuvent être utilisés pour décrire la courbe de Wöhler, dont la loi de Basquin présentée ici,

$$S = S'_f N^m, \quad (1.5)$$

où S'_f est le coefficient de résistance à la fatigue et m est l'exposant de résistance à la fatigue. Cette expression sous forme logarithmique devient

$$\log(N) = \log(a) - b \log(S), \quad (1.6)$$

avec $a = S'_f{}^{-\frac{1}{m}}$ et $b = \frac{1}{m}$ la pente de la courbe de Wöhler.

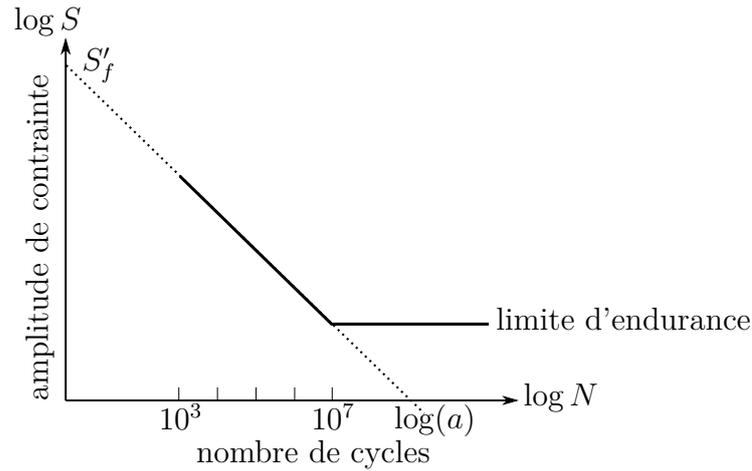


FIGURE 1.15 – Courbe de Wöhler représentée sur une échelle logarithmique

1.2.3 Modèle probabiliste de la courbe de Wöhler

Une revue des différents modèles probabilistes de la courbe de Wöhler pour les matériaux métalliques et composites est proposée par Barbosa et al. [27]. Le plus populaire et le plus adapté au contexte industriel est présenté ici. Ce modèle suppose que la variance de $\log(N)$ est constante sur la plage considérée, sur la base de l'hypothèse de données homoscédastiques [28]. Ceci peut être assuré en prenant le paramètre a de la loi de Basquin dans l'Équation (1.6) comme étant une variable aléatoire et en considérant la pente b comme un coefficient déterministe. Dans le cas général, les paramètres a et b devraient tous deux être modélisés par des variables aléatoires, mais la variabilité du paramètre b n'est pas significative selon les données expérimentales [28]. Par conséquent, le paramètre b est ici considéré comme une valeur déterministe.

Le nombre de cycles jusqu'à rupture est considéré comme une variable aléatoire décrite par une loi Log-normale ou de Weibull. De nombreuses comparaisons entre ces deux lois ont été réalisées [28, 29, 30]. La loi Log-normale est adoptée ici pour représenter la courbe de Wöhler probabiliste. Ce choix est cohérent avec le choix d'une distribution Log-normale pour la durée de vie proposé par les experts dans l'entreprise. Ainsi, le paramètre a dans l'équation (1.6) est une variable aléatoire supposée suivre la distribution Log-normale $a \sim \mathcal{LN}(\mu_a, \sigma_a)$. La pente b de la courbe de Wöhler peut être considérée comme une constante. Le modèle probabiliste de la courbe de Wöhler est présenté sur la Figure 1.16.

Il est mentionné dans les travaux de Hanaki et al. [31] que le niveau de dispersion du nombre de cycles devrait diminuer lorsque l'amplitude de la contrainte augmente, plusieurs modèles plus complexes sont ainsi proposés. Ce phénomène est ici pris en considération. La variance de $\log(a)$ est constante pour différents niveaux d'amplitude de contrainte comme $\log(a) \sim \mathcal{N}(\mu_a, \sigma_a)$, mais la variance du paramètre a diminue avec l'augmentation de l'amplitude de contrainte, ainsi que la variance du nombre de cycles.

Pour déterminer les paramètres de la courbe de Wöhler probabiliste (courbe P-S-N) par des tests d'éprouvettes, de nombreux travaux sont proposés dans la littérature. Li et al. [32] donnent une revue concernant la détermination de la courbe P-S-N en fatigue. Trois types de méthodes [33] sont distinguées : la méthode de groupe basée sur une relation li-

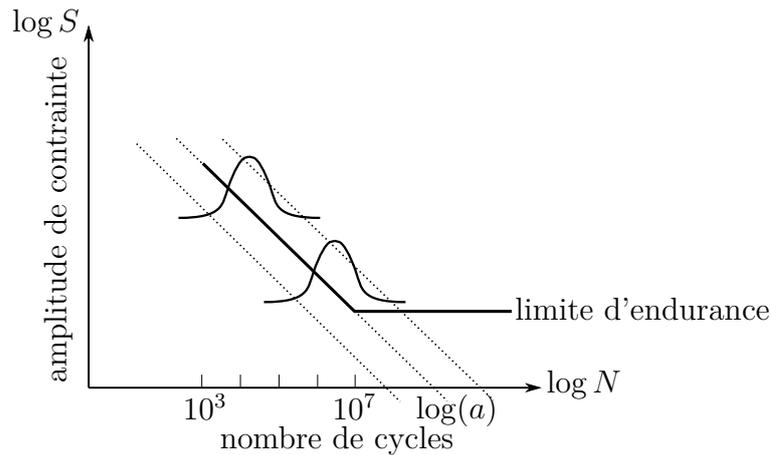


FIGURE 1.16 – Courbe de Wöhler probabiliste (courbe P-S-N) représentée sur une échelle logarithmique

néaire entre $\log S$ et $\log N$, l'approche de la distribution de Weibull et l'analyse statistique avancée. Pour cette dernière, on peut citer la méthode du maximum de vraisemblance et la méthode de Monte Carlo [34], ou encore l'algorithme différentiel d'anti-stigmergie [35]. Hanaki et al. [31] proposent un modèle de courbe P-S-N basé sur la distribution de la résistance à la fatigue. Pavlou [36] présente l'étude de la courbe P-S-N combinée à la loi d'accumulation des dommages. De plus, de nombreux chercheurs traitent de la réduction du nombre d'éprouvettes pour la détermination des courbes P-S-N [37, 29]. Certaines méthodes pratiques sont présentées dans les normes ASTM [38], Eurocode [39] et la norme de la société japonaise des matériaux [40]. Ces normes sont largement adoptées dans le monde industriel mécanique. Cependant, les essais expérimentaux sur une structure complexe conduisent à des courbes de Wöhler significativement différentes par rapport aux essais sur éprouvettes. D'après Petinov et al. [41], le comportement en fatigue des structures et des éprouvettes sont compatibles dans la phase d'initiation de la croissance consécutive des fissures, mais présentent une inadéquation sérieuse dans la phase de propagation des fissures, dépendant de la géométrie, de la contrainte résiduelle des soudures et des chargements. Dans l'étude de Sonsino et al. [4] pour un arbre d'entraînement d'une boîte de vitesse automatique sous couple opérationnel, il est montré que la courbe P-S-N obtenue avec des éprouvettes non entaillées et une charge axiale ne peut pas bien reproduire le comportement en fatigue des composants. Certains travaux utilisent des facteurs déterminés empiriquement pour représenter la différence entre les éprouvettes à petite échelle et les composants à échelle réelle [42]. Il est donc recommandé d'utiliser la courbe de Wöhler pour des composants à l'échelle réelle afin d'évaluer la durée de vie en fatigue des cordons de soudure de la structure.

1.3 Conclusion

Sur la base du contexte industriel et d'un état de l'art sur la fatigue des matériaux proposés dans ce premier chapitre, le manuscrit est composé des chapitres 2 et 3 représentatifs de la contribution scientifique principale dans cette thèse. Le chapitre 1 est concentré

à l'application des travaux de thèse au contexte industriel. L'objectif de la thèse est de construire un dialogue essai-calcul (Figure 1.17) pour mieux corrélérer les résultats issus de l'essai et de la simulation numérique. A cet effet, le chapitre 2 propose tout d'abord le traitement et l'analyse des données groupées expérimentales mesurées dans l'essai synthétique de l'essieu souple, cette contribution est représentée par le rectangle A sur la figure. Dans le but d'évaluer la variabilité observée dans les essais expérimentaux, différentes méthodes d'estimation sont également introduites. La seconde contribution est proposée dans le chapitre 3, elle concerne l'identification des incertitudes liées à la courbe de Wöhler afin que la durée de vie issue du calcul par éléments finis puisse représenter la variabilité estimée avec la durée de vie expérimentale. Cette approche d'identification, représentée par le rectangle B sur la figure, est basée sur un problème inverse.

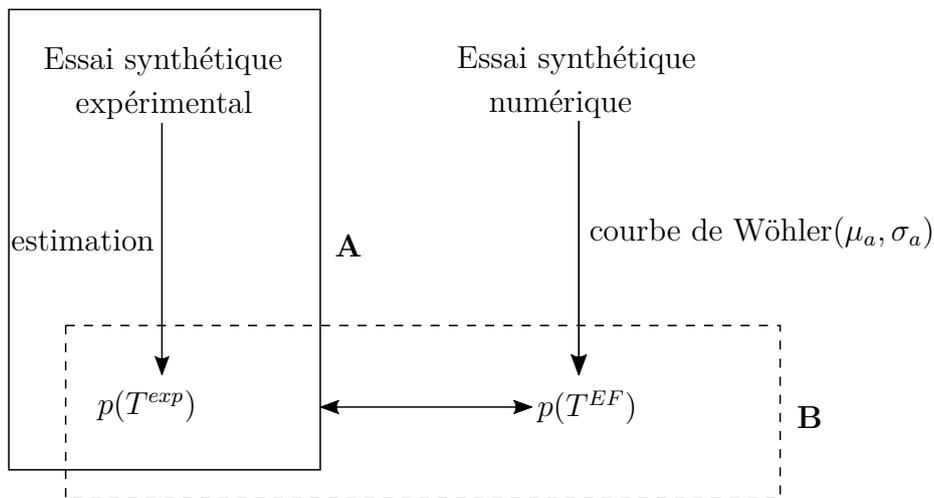


FIGURE 1.17 – Dialogue essai-calcul

Chapitre 2

Estimation of grouped data

In this chapter, the grouped data of experimental fatigue lifetime are estimated with different methods. The estimated probability density, reliability and failure rate are obtained and discussed.

Contents

2.1	Introduction	25
2.2	Presentation of experimental grouped data	26
2.2.1	Homogeneous grouped data	26
2.2.2	Non-homogeneous grouped data	26
2.3	MLE for Weibull and Log-normal distributions	27
2.3.1	Method and analysis	27
2.3.1.1	Maximum likelihood estimation	27
2.3.1.2	Fisher information and confidence interval	28
2.3.1.3	Reliability analysis	30
2.3.2	Weibull distribution	31
2.3.2.1	Homogeneous grouped data	32
2.3.2.2	Non-homogeneous grouped data	33
2.3.3	Log-normal distribution	34
2.3.3.1	Homogeneous grouped data	34
2.3.3.2	Non-homogeneous grouped data	35
2.3.4	Comparison of estimations obtained by Weibull and Log-normal distributions	36
2.3.4.1	Homogeneous grouped data	36
2.3.4.2	Non-homogeneous grouped data	37
2.4	Goodness of fit tests for grouped data	39
2.4.1	Kolmogorov-Smirnov test for grouped data	39
2.4.1.1	Kolmogorov-Smirnov test for continuous data	39
2.4.1.2	Conover's procedure	40
2.4.1.3	Result for simulated data of Log-normal distribution	41

2.4.1.4	Result for industrial experimental data of Log-normal distribution	42
2.5	Bayesian estimation for Weibull distribution and Log-normal distribution	43
2.5.1	Method of Bayesian estimation	44
2.5.1.1	Monte Carlo method (importance sampling)	44
2.5.1.2	Markov chain Monte Carlo	44
2.5.2	Bayesian estimation for Weibull distribution	47
2.5.2.1	Bayesian estimation of one parameter	48
2.5.2.2	Bayesian estimation of two parameters	49
2.5.3	Bayesian estimation for Log-normal distribution	51
2.5.3.1	Bayesian estimation of one parameter	51
2.5.3.2	Bayesian estimation of two parameters	54
2.5.3.3	Calculation of the posterior estimator by MCMC	54
2.5.4	Comparison among the different Bayesian estimation methods	55
2.6	Estimation with multinomial distribution for grouped data	57
2.6.1	Definition	57
2.6.2	Conjugate prior and posterior distribution	58
2.6.2.1	Determination of prior distribution parameters value	59
2.6.2.2	Results	60
2.7	Simulated data estimation results	60
2.7.1	Simulating method and simulated samples	61
2.7.2	Bayesian estimation for Log-normal distribution with grouped data	63
2.7.2.1	Normal-inverse-gamma prior	63
2.7.2.2	Uniform prior	64
2.7.2.3	Bayesian estimation for multinomial distribution	64
2.8	Conclusion	65

2.1 Introduction

In the Renault company, the experimental tests of endurance are done for different levels of structure. This thesis focuses on the endurance tests of chassis components. In those experimental tests, the interest is the location of cracks and the duration from the start of the test to first macro crack. Usually the cracks appear near the welded joints of the sheet metal part assembly. Thus the time duration is also called the lifetime of each welded joint. The fatigue lifetime is here used for mechanical fatigue failure.

The endurance tests are performed for several theoretically identical structures of the same design. The observed dispersion of fatigue lifetime is due to micro cracks in the material. Estimation methods are used to study and describe the dispersion phenomenon. But, as the costs of full-scale component fatigue tests are generally high [43], the number of tests is often small and the observation data are highly-censored. As mentioned in section 1.1.5, the measurement of cracks by dye penetrant testing leads to interval-censored data (grouped data) for fatigue lifetime.

In general, the fatigue lifetime is supposed to follow the Weibull distribution or the Log-normal distribution. There are many discussions about the choice between those two distributions [28]. In the work of Schijve [2], the two distributions are separately estimated from two series of experimental data for sample size equal to 15 and 30. The results show that both of distributions are well-fitted. In this present work, the Log-normal distribution is chosen according to the experts opinion of the company.

Several works deal with the grouped data estimation. For Log-normal distribution, Nishino and Kakamu [44] proposed an estimator of the parameter explicitly based on selected order statistics, which requires less calculation than maximum likelihood estimation. Migdadi and Al-Batah [45] realized a Bayesian estimation for one parameter of Log-normal distribution with grouped data. For Weibull distribution, Guure et al. [46] proposed a work on Bayesian estimation of Weibull distribution with Gamma prior distribution. Sadek and Alhabi [47] used the ranked set sampling data which contain more information compared to simple random sampling to do the Bayesian estimation of Weibull distribution. Kundu and Mitra [48] discussed the different prior distributions when the shape parameter is known or unknown for Weibull distribution based on left truncated and right censored data.

This chapter deals with estimation of grouped data. Two estimation methods are used : the maximum likelihood estimation (MLE) and the Bayesian method. The MLE method considers the parameter as a deterministic value and the Bayesian estimation takes the parameter as a random variable. The continuous Log-normal and Weibull distributions, as well as the discrete multinomial distribution are proposed to describe the variability of grouped data of fatigue lifetime. In the end, the proposed estimation methods are tested for simulated data to validate their performances. This chapter corresponds to the part A in the Figure 1.17 describing the dialogue between experimental test, and numerical simulation.

2.2 Presentation of experimental grouped data

2.2.1 Homogeneous grouped data

As mentioned in section 1.1.5, the technique of observation during experimental tests induces grouped data for fatigue life. Grouped data means that the exact failure time is not measured, only the number of failures in a time interval is observed. For each welding joint, the experimental data of fatigue lifetime are in the form of grouped data, as shown in Figure 2.1. The overall time-frame is divided into $m+1$ intervals $[t_{j-1}, t_j]$ by the inspection time with $j \in \llbracket 1, m+1 \rrbracket$. The initial and end time of the test are respectively $t_0 = 0$ and t_m . The interval $[t_m, t_{m+1}]$ with $t_{m+1} = \infty$ denotes the time after the end of test and t_j with $j \in \llbracket 1, m \rrbracket$ are the instants of dye penetrant inspection. The number of failures in each time interval $[t_{j-1}, t_j]$ is a random variable N_j^f . The random vector $\underline{N}^f = (N_1^f, \dots, N_m^f, N_{m+1}^f)$ represents the history of failures over the entire time-frame. The observed experimental data n_j^f can be considered as a realization of the random variable of number of failures N_j^f , with the total number of experimental observation number $n^f = \sum_{j=1}^{m+1} n_j^f$. The number N_{m+1}^f in the $m+1^{\text{th}}$ interval corresponds to the number of experiments with no failure detected during the test for the considered welding joint.

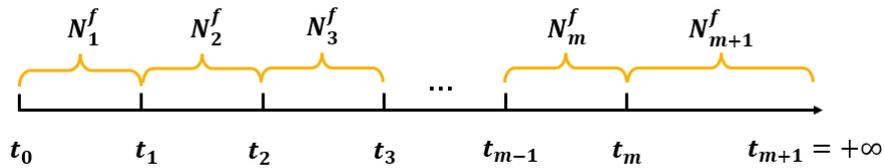


FIGURE 2.1 – Grouped data

For the different experimental tests, when the observations lead to the same interval of grouped data, then they can be grouped together and are called homogeneous grouped data. Nevertheless, if the observation interval is different from one test to another, non-homogeneous grouped data are generated.

2.2.2 Non-homogeneous grouped data

Non-homogeneous grouped data are effectively induced when the observation intervals are different for each test performed on theoretically identical structures. Suppose there are l times experimental tests, for the i^{th} test, m_i observation intervals are considered with $i \in \llbracket 1, l \rrbracket$. The number of crack initiations $N_{i,j}^f$ in each time interval $[t_{i,j-1}, t_{i,j}]$ is considered as a random variable with $j \in \llbracket 1, m_i + 1 \rrbracket$, $t_0 = 0$ and $t_{m_i+1} = +\infty$. Figure 2.2 shows an example of non-homogeneous grouped data, where $m_1 = 5$ for the test $i = 1$, $m_2 = 4$ for the test $i = 2$ and m_l for the test $i = l$. Each test has different observation intervals.

In practice, Renault experimental endurance tests lead to homogeneous or non-homogeneous grouped data, depending on the dye penetrant inspection which checks

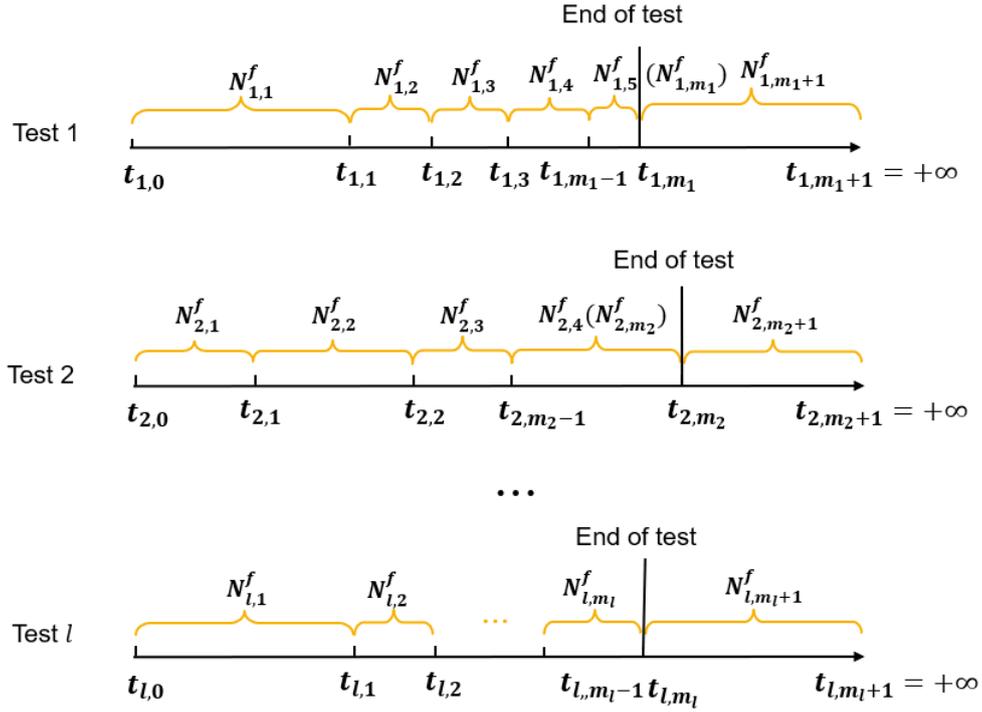


FIGURE 2.2 – Non-homogeneous grouped data

surface-breaking defects during the test. In this chapter, different methods are now proposed to estimate the distribution of these experimental data based on different hypothesis.

2.3 MLE for Weibull and Log-normal distributions

2.3.1 Method and analysis

The fatigue lifetime T of a welded joint is here modeled by a random variable characterized by a parameter θ . Let us denote the cumulative distribution function (CDF) by $F(t|\theta)$, the probability density function (PDF) by $f(t|\theta)$ and the probability mass function (PMF) $P_\theta(t)$ in the discrete case.

2.3.1.1 Maximum likelihood estimation

Maximum likelihood estimation (MLE) is here considered to estimate the parameters of a probability distribution. For homogeneous grouped data shown in Figure 2.1, N_j^f is the number of crack initiation in each time interval $[t_{j-1}, t_j]$ with $j = 1, \dots, m + 1$. In this case, the realisation of random vector $\underline{N}^f = (N_1^f, \dots, N_m^f, N_{m+1}^f)$ is $\underline{n}^f = (n_1^f, \dots, n_m^f, n_{m+1}^f)$, also called observations. The maximum likelihood function $\mathcal{L}(\theta|\underline{N}^f)$ of θ with observation \underline{n}^f is

$$\mathcal{L}(\theta|\underline{n}^f) = \prod_{j=1}^{m+1} [F(t_j|\theta) - F(t_{j-1}|\theta)]^{n_j^f}, \quad (2.1)$$

where F is the CDF parametrized by θ , m denotes the number of intervals, n_j^f is the realisation of the number of failures in j -th interval and $n^f = \sum_{i=1}^{m+1} n_i^f$ is the total number of experimental observations. The likelihood function is the joint probability distribution of the sample, but seen as a function of parameters. It measures the goodness of fit of a statistical model to a sample of data. The MLE is used to estimate parameters of probability distribution by maximizing a likelihood function, so that the observed sample is the most probable with those parameters for the assumed statistical model.

For non-homogeneous grouped data, as in Figure 2.2, with the crack initiation number $N_{i,j}^f$ in the j^{th} time interval of i^{th} experimental test, and the observations $\underline{n}^f = (n_{1,1}^f, \dots, n_{1,m_1}^f, n_{1,m_1+1}^f, n_{2,1}^f, \dots, n_{2,m_2}^f, n_{2,m_2+1}^f, n_{l,1}^f, \dots, n_{l,m_l}^f, n_{l,m_l+1}^f)$, the maximum likelihood function is

$$\mathcal{L}(\theta|\underline{n}^f) = \prod_{i=1}^l \prod_{j=1}^{m_i+1} [F(t_{i,j}|\theta) - F(t_{i,j-1}|\theta)]^{n_{i,j}^f}. \quad (2.2)$$

The maximum likelihood estimators are the parameters which maximize the likelihood function

$$\hat{\theta} = \arg \max_{\theta} \log \mathcal{L}(\theta|\underline{n}^f), \quad (2.3)$$

where $\hat{\theta}$ is the estimator of θ . This optimization problem can be solved by an iterative scheme, such as Newton-Raphson method.

2.3.1.2 Fisher information and confidence interval

The Fisher information quantifies the amount of information about the parameter θ in a sufficient statistic and data, Fisher introduced the following measure according to [49]:

The Fisher information $I(\theta)$ of a r.v. X in state space \mathcal{X} about θ is defined as

$$I(\theta) = \begin{cases} \sum_{x \in \mathcal{X}} \left(\frac{d}{d\theta} \log f(x|\theta) \right)^2 p_{\theta}(x), & \text{if } X \text{ is discrete} \\ \int_{\mathcal{X}} \left(\frac{d}{d\theta} \log f(x|\theta) \right)^2 p_{\theta}(x) dx, & \text{if } X \text{ is continuous} \end{cases} \quad (2.4)$$

where the derivation $\frac{d}{d\theta} \log f(x|\theta)$ is the score function.

The Fisher information is important in statistical applications. In the frequentist paradigm, Fisher information is used to construct hypothesis tests and the confidence intervals of maximum likelihood estimators. In the Bayesian paradigm, Fisher information is used to define a non-informative Jeffrey prior distribution [49]. In this subsection, the calculation of confidence interval for maximum likelihood estimator from grouped data is presented.

It is noticed that the Fisher information for grouped data is different from continuous data. Aludaat et al. [50] give the MLE and Bayesian estimations of Burr Type X distribution for grouped data, whose Fisher information is presented. And Xia et al. [51] prove

that the maximum likelihood estimators from grouped data of Log-normal distribution is uniquely existed. In what follows, the calculation of Fisher information for grouped data is detailed.

By the definition, the Fisher information is

$$I(\theta) = \mathbb{E} \left[\left(\frac{\partial \log f(x|\theta)}{\partial \theta} \right)^2 \right], \quad (2.5)$$

for notational simplicity, f will be used in the place of $f(x|\theta)$.

Under mild regular condition [49], the Fisher information can be written as

$$I(\theta) = \mathbb{E} \left[\left(\frac{\partial \log f}{\partial \theta} \right)^2 \right] = Var \left[\frac{\partial \log f}{\partial \theta} \right] + \left(\mathbb{E} \left[\frac{\partial \log f}{\partial \theta} \right] \right)^2, \quad (2.6)$$

where $\left(\mathbb{E} \left[\frac{\partial \log f}{\partial \theta} \right] \right)^2$ can be simplified to zero.

Then the Equation (2.6) becomes

$$I(\theta) = Var \left[\frac{\partial \log f}{\partial \theta} \right], \quad (2.7)$$

which shows that the Fisher information represents the variance of score function.

By a straightforward calculation, we can obtain

$$I(\theta) = -\mathbb{E} \left[\frac{\partial^2 \log f}{\partial \theta^2} \right]. \quad (2.8)$$

For n independent and identically distributed observations, the PDF $f(x|\theta)$ is replaced by the likelihood function $\mathcal{L}(x|\theta) = \prod_{i=1}^n f(x_i|\theta)$, and the Fisher information is n times the unit Fisher information $I(\theta)$

$$I_n(\theta) = -\mathbb{E} \left[\frac{\partial^2 \log \mathcal{L}}{\partial \theta^2} \right] = nI(\theta). \quad (2.9)$$

In the case of grouped data, the likelihood function is given by Equation (2.1). Suppose the probability of fatigue lifetime in interval $[t_{-1}, t_j]$ is $P_j = F(t_j|\theta) - F(t_{-1}|\theta)$ with $\underline{\theta} = (\theta_1, \theta_2)$, the first and second deviations of a function with respect to a vector equals to the Jacobian and Hessian matrix respectively, thus the first and second deviations of $\log \mathcal{L}$ with respect to $\underline{\theta}$ are calculated

$$\frac{\partial \log \mathcal{L}}{\partial \underline{\theta}} = \begin{bmatrix} \frac{\partial \log \mathcal{L}}{\partial \theta_1} & \frac{\partial \log \mathcal{L}}{\partial \theta_2} \end{bmatrix} = \begin{bmatrix} \sum_{j=1}^{m+1} \frac{N_j^f}{P_j} \frac{\partial P_j}{\partial \theta_1} & \sum_{j=1}^{m+1} \frac{N_j^f}{P_j} \frac{\partial P_j}{\partial \theta_2} \end{bmatrix}, \quad (2.10)$$

$$\begin{aligned} \frac{\partial^2 \log \mathcal{L}}{\partial \underline{\theta}^2} &= \begin{bmatrix} \frac{\partial^2 \log \mathcal{L}}{\partial \theta_1^2} & \frac{\partial^2 \log \mathcal{L}}{\partial \theta_1 \partial \theta_2} \\ \frac{\partial^2 \log \mathcal{L}}{\partial \theta_1 \partial \theta_2} & \frac{\partial^2 \log \mathcal{L}}{\partial \theta_2^2} \end{bmatrix} \\ &= \begin{bmatrix} \sum_{j=1}^{m+1} \frac{N_j^f}{P_j^2} \left(\frac{\partial P_j}{\partial \theta_1} \right)^2 - \frac{N_j^f}{P_j} \frac{\partial^2 P_j}{\partial \theta_1^2} & \sum_{j=1}^{m+1} \frac{N_j^f}{P_j^2} \frac{\partial P_j}{\partial \theta_1} \frac{\partial P_j}{\partial \theta_2} - \frac{N_j^f}{P_j} \frac{\partial^2 P_j}{\partial \theta_1 \partial \theta_2} \\ \sum_{j=1}^{m+1} \frac{N_j^f}{P_j^2} \frac{\partial P_j}{\partial \theta_1} \frac{\partial P_j}{\partial \theta_2} - \frac{N_j^f}{P_j} \frac{\partial^2 P_j}{\partial \theta_1 \partial \theta_2} & \sum_{j=1}^{m+1} \frac{N_j^f}{P_j^2} \left(\frac{\partial P_j}{\partial \theta_2} \right)^2 - \frac{N_j^f}{P_j} \frac{\partial^2 P_j}{\partial \theta_2^2} \end{bmatrix}. \end{aligned} \quad (2.11)$$

For grouped data, the observations are the numbers of failure N_j^f in each time interval $[t_{j-1}, t_j]$ with $j = 1, \dots, m+1$, the exact fatigue lifetime t has not been identified, therefore the Fisher information is the expectation with respect to N_j^f . Suppose the data size is n , then the expectation of fatigue lifetime in j^{th} interval is $\mathbb{E}(N_j^f) = nP_j$, the Fisher matrix is

$$-\mathbb{E}\left(\frac{\partial^2 \log \mathcal{L}}{\partial \theta^2}\right) = n \begin{bmatrix} \sum_{j=1}^{m+1} \frac{1}{P_j} \left(\frac{\partial P_j}{\partial \theta_1}\right)^2 - \frac{\partial^2 P_j}{\partial^2 \theta_1} & \sum_{j=1}^{m+1} \frac{1}{P_j} \frac{\partial P_j}{\partial \theta_1} \frac{\partial P_j}{\partial \theta_2} - \frac{\partial^2 P_j}{\partial \theta_1 \partial \theta_2} \\ \sum_{j=1}^{m+1} \frac{1}{P_j} \frac{\partial P_j}{\partial \theta_1} \frac{\partial P_j}{\partial \theta_2} - \frac{\partial^2 P_j}{\partial \theta_1 \partial \theta_2} & \sum_{j=1}^{m+1} \frac{1}{P_j} \left(\frac{\partial P_j}{\partial \theta_2}\right)^2 - \frac{\partial^2 P_j}{\partial \theta_2^2} \end{bmatrix}. \quad (2.12)$$

Some numerical calculation problems may be caused by the observation time where $t_0 = 0$ and $t_{k+1} = +\infty$ of the grouped data, hence $P_1 = F(t_i; \theta)$ and $P_{k+1} = 1 - F(t_k)$ need to be dealt with differently (see appendix A).

According to asymptotic property of maximum likelihood estimator $\hat{\theta}$, the difference between the true parameter value θ^* and the maximum likelihood estimator converges in distribution to a normal distribution [49],

$$\sqrt{n}(\hat{\theta} - \theta^*) \xrightarrow{D} \mathcal{N}(0, I^{-1}(\theta^*)), \quad \text{as } n \rightarrow \infty. \quad (2.13)$$

For n large enough, the difference is approximately normally distributed

$$(\hat{\theta} - \theta^*) \stackrel{D}{\approx} \mathcal{N}\left(0, \frac{1}{I_n(\theta^*)}\right). \quad (2.14)$$

It shows that the estimator $\hat{\theta}$ is close to the true value θ^* with a normal distribution whose standard deviation is the inverse of square root of the Fisher information [49]. Thus larger data size n reduces the standard deviation bringing a potential closer $\hat{\theta}$ to θ^* . Relatively, the true value θ^* is around the estimator $\hat{\theta}$ in this normal distribution, thus with this asymptotic property, the confidence interval at level α interval is obtained

$$\left(\hat{\theta} - u_{1-\frac{\alpha}{2}} \sqrt{I_n^{-1}(\hat{\theta})}, \hat{\theta} + u_{\frac{\alpha}{2}} \sqrt{I_n^{-1}(\hat{\theta})}\right), \quad (2.15)$$

where $u_{1-\frac{\alpha}{2}}$ is the $1 - \frac{\alpha}{2}$ fractile of standard normal distribution. For example, in the case of 95% confidence level, $u_{\frac{\alpha}{2}} = 1 - u_{1-\frac{\alpha}{2}} = 1.96$.

2.3.1.3 Reliability analysis

When the fatigue lifetime distribution of welded joint is estimated, the reliability analysis can be done.

Considering the fatigue lifetime T is a positive real *r.v.* with PDF $f(t)$ and CDF $F(t) = P(X \leq t)$, the reliability $R(t)$ describes the ability of a system or component to function under stated conditions for a specified period of time [52]. Theoretically it is defined as the probability of success at time t with $t \geq 0$

$$R(t) = \bar{F}(t) = 1 - F(t) = \int_t^{\infty} f(u) du, \quad (2.16)$$

where $\bar{F}(t)$ is the complement of $F(t)$.

The failure rate $\lambda(t)$ is the limit of the probability that a failure occurs in the time interval $[t, t + \delta t]$, given that no failure has occurred before time t [53], it is defined as

$$\lambda(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t < T \leq t + \Delta t | T > t)}{\Delta t}. \quad (2.17)$$

When the *r.v.* T is absolutely continuous, the failure rate is calculated by

$$\lambda(t) = \frac{1}{\bar{F}(t)} \frac{d}{dt} F(t) = \frac{f(t)}{\bar{F}(t)} = \frac{f(t)}{R(t)}. \quad (2.18)$$

The cumulative failure rate is given by

$$\Lambda(t) = \int_0^t \lambda(u) du. \quad (2.19)$$

The Mean Time To Failure (MTTF) is the average time before the failure occurs which is useful in application, it can be calculated as follow

$$MTTF = \mathbb{E}[X] = \int_0^{\infty} t dF(t) = \int_0^{\infty} t f(t) dt. \quad (2.20)$$

2.3.2 Weibull distribution

For three-parameters Weibull distribution, the cumulative distribution function is

$$F = 1 - e^{-\left(\frac{t-c}{\alpha}\right)^\beta}, \quad (2.21)$$

where α is the scale parameter, β the shape parameter and c the location parameter.

We use the three-parameter Weibull distribution to describe the lifetime when there is a guaranteed period c without failure. In our case it is not suitable, because with the experience on the fatigue lifetime observations there is not a fixed period without failure. Thus two-parameters, Weibull distribution, with $c = 0$, is proposed in this case, whose CDF is

$$F(t|\alpha, \beta) = 1 - e^{-\left(\frac{t}{\alpha}\right)^\beta}, \quad t \geq 0, \quad (2.22)$$

or

$$F(t|\gamma, \beta) = 1 - e^{-\frac{t^\beta}{\gamma}}, \quad t \geq 0, \quad (2.23)$$

where $\gamma = \alpha^\beta$.

So its likelihood function for grouped data with Weibull distribution is

$$\mathcal{L}(\alpha, \beta | \underline{N}^f) = \prod_{j=1}^{m+1} \left[e^{-\left(\frac{t_{j-1}}{\alpha}\right)^\beta} - e^{-\left(\frac{t_j}{\alpha}\right)^\beta} \right]^{N_j^f}, \quad (2.24)$$

or

$$\mathcal{L}(\gamma, \beta | \underline{N}^f) = \prod_{j=1}^{m+1} \left[e^{-\frac{t_{j-1}^\beta}{\gamma}} - e^{-\frac{t_j^\beta}{\gamma}} \right]^{N_j^f}. \quad (2.25)$$

The estimators are calculated by maximizing numerically the likelihood function with Newton-Raphson method.

Some statistical indicators such as mean value μ_W , standard deviation σ_W , median M_W and percentile are used to make a decision in industry. The calculation for some of those indicators for Weibull distribution is here given :

$$\begin{cases} \mu_W = \alpha\Gamma(1 + \frac{1}{\beta}), \\ \sigma_W = \alpha \left[\Gamma(1 + \frac{2}{\beta}) - \Gamma^2(1 + \frac{1}{\beta}) \right]^{\frac{1}{2}}, \\ M_W = \alpha(\log 2)^{(1/\beta)}, \end{cases} \quad (2.26)$$

where Γ is the gamma function $\Gamma(z) = \int_0^\infty x^{z-1}e^{-x} dx$.

The formulation of the Fisher information for Weibull distribution in the case of grouped data is expressed in Equation (2.12). The confidence interval of estimators can be calculated from the Fisher information.

The estimation for Weibull distribution is done for homogeneous grouped data at first, and then for non-homogeneous grouped data. In our study, the unit of fatigue lifetime is considered as a percentage of the nominal time.

2.3.2.1 Homogeneous grouped data

For the real Renault experimental fatigue lifetime grouped data for a welded joint presented in Figure 2.3, the estimation by MLE method for Weibull distribution is done. The estimated Weibull probability density is presented and compared with Log-normal distribution later in Figure 2.4 in section 2.3.4.1. The result shows that the mean value μ_W is equal to 115.78%, the variance is equal to 0.1122, the standard deviation σ_W is equal to 34.96%, the median M_W is equal to 116.18% and the 10 percentile is equal to 69.70%.

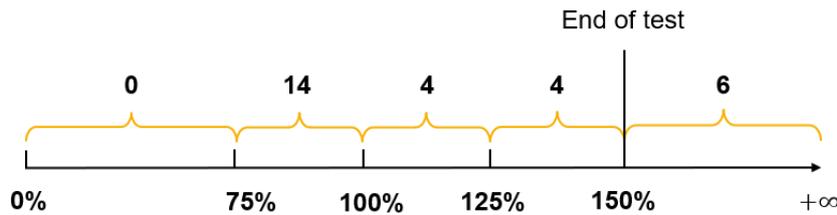


FIGURE 2.3 – Example of Renault grouped data

2.3.2.1.1 Fisher information and confidence interval

The Fisher matrix in Equation (2.12) could be written as

$$I_n(\theta) = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}. \quad (2.27)$$

The CDF of two-parameters Weibull distribution is substituted into the Equation (2.12), the details of calculation can be found in appendix A. The terms A_{11} and A_{22} are expressed as

$$A_{11} = \sum_{j=0}^{m+1} \frac{1}{P_j} \left(\frac{\partial P_j}{\partial \alpha} \right)^2 - \frac{\partial^2 P_j}{\partial \alpha^2} \quad (2.28)$$

$$A_{22} = \sum_{j=0}^{m+1} \frac{1}{P_j} \left(\frac{\partial P_j}{\partial \beta} \right)^2 - \frac{\partial^2 P_j}{\partial \beta^2} \quad (2.29)$$

With the data presented in Figure 2.3, the maximum likelihood estimators are calculated. The obtained values are $\hat{\alpha} = 1.2832$ and $\hat{\beta} = 3.6875$, for their 95% confidence interval [1.0984, 1.4679] and [2.0652, 5.3098] respectively.

2.3.2.1.2 Reliability analysis

With two-parameters Weibull distribution, the reliability is expressed as

$$R(t) = \exp \left[- \left(\frac{t}{\alpha} \right)^\beta \right] \quad (2.30)$$

The failure rate is

$$\lambda(t) = \frac{\beta}{\alpha} \left(\frac{t}{\alpha} \right)^{\beta-1}. \quad (2.31)$$

With the distribution estimated above, the Mean Time to Failure is equal to 115.78%.

The Weibull distribution is often used in the reliability analysis. The shape parameter β has marked effects on the behavior of distribution. For the failure rate $\lambda(t)$, when $\beta < 1$, the failure rate decreases with time. For $\beta = 1$, the Weibull distribution has an almost constant failure rate. When $\beta > 1$, the failure rate increases with time (see Figure 2.5(b)). It means more components will fail in a unit of time with time evolution.

2.3.2.2 Non-homogeneous grouped data

An example of data for 34 experimental tests is here chosen, there are 16 different observation intervals, the raw data can be found in appendix B. The maximum likelihood estimators are $\hat{\alpha} = 2.6651$ and $\hat{\beta} = 1.0733$. The estimated PDF is presented and compared with Log-normal distribution law later in Figure 2.6. The mean value is equal to 243.93%, the variance is equal to 5.8456, the standard deviation is equal to 241.78%, the median is equal to 178.16% and the 10-percentile is equal to 30.80%. The estimated MTTF is equal to 243.93%.

2.3.2.2.1 Fisher information and confidence interval

For non-homogeneous grouped data, the Equation (2.9) is no longer valid. The Fisher information $I_n(\theta)$ is calculated from the sum of $I(\theta)$ for each observation

$$I_n(\theta) = I^{(1)}(\theta) + I^{(2)}(\theta) + \dots + I^{(n)}(\theta), \quad (2.32)$$

where n is the total number of observations.

With the non-homogeneous grouped data example, the confidence intervals of maximum likelihood estimators are obtained : [0.3575, 4.6562] for α and [0.5946, 1.5519] for β .

2.3.2.2.2 Reliability analysis

For Weibull distribution, Equation (2.30) and Equation (2.31) are used for reliability analysis.

With the shape parameter β close to 1, the failure rate increases with time at first stage and then stay almost constant (see Figure 2.7). Compared to the results obtained with homogenous grouped data, the reliability and failure rate are very different. It means that the reliability performance can depend on welding joints and rear axle projects.

2.3.3 Log-normal distribution

The likelihood function for grouped data is given by

$$\mathcal{L}(\theta|\underline{N}^f) = \prod_{j=1}^{m+1} [F(t_j|\theta) - F(t_{j-1}|\theta)]^{N_j^f}, \quad (2.33)$$

where F is the CDF of Log-normal distribution.

For Log-normal distribution, the CDF is expressed as

$$F(t) = \frac{1}{2} + \frac{1}{2} \operatorname{erf} \left[\frac{\log(t) - \mu}{\sigma\sqrt{2}} \right], \quad (2.34)$$

where the error function is $\operatorname{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-t^2} dt$. The time t is the realisation of the fatigue lifetime random variable T . If a *r.v.* T follows a Log-normal distribution, then the natural logarithm of T is normally distributed, $\log(T) \sim \mathcal{N}(\mu, \sigma^2)$ with mean μ and variance σ . Now let $\delta = (\mu_T, \sigma_T)$ represent the distribution parameters of random variable T , it will be estimated by the maximum likelihood method.

The corresponding likelihood function for grouped data is

$$\mathcal{L}(\mu, \sigma|\underline{N}^f) = \prod_{j=1}^{m+1} \left[\frac{1}{2} \operatorname{erf} \left[\frac{\log(t_j) - \mu}{\sigma\sqrt{2}} \right] - \frac{1}{2} \operatorname{erf} \left[\frac{\log(t_{j-1}) - \mu}{\sigma\sqrt{2}} \right] \right]^{N_j^f}. \quad (2.35)$$

The mean value μ_L , standard deviation σ_L and median M_L of the Log-normal distribution are

$$\begin{cases} \mu_L = e^{\mu + \frac{\sigma^2}{2}}, \\ \sigma_L = [(e^{\sigma^2} - 1)e^{2\mu + \sigma^2}]^{\frac{1}{2}}, \\ M_L = e^{\mu}. \end{cases} \quad (2.36)$$

2.3.3.1 Homogeneous grouped data

The estimation of Log-normal distribution is done with homogeneous grouped data at first, and then for non-homogeneous grouped data.

With the data set presented in Figure 2.3, the estimated parameters of Log-normal distribution are $\hat{\mu} = 0.1178$ and $\hat{\sigma} = 0.2887$. The mean value is equal to 117.29%, the variance is equal to 0.1195, the standard deviation is equal to 34.57%, the median is equal to 112.50%, and the 10 percentile is equal to 77.72%.

2.3.3.1.1 Fisher information and confidence interval

The Fisher matrix in Equation (2.12) could be written as

$$I_n(\theta) = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}. \quad (2.37)$$

For the simplicity of computation, a log transformation is done with experimental fatigue lifetime data to follow the normal distribution. According to the Log-normal distribution property, $T \sim \mathcal{LN}(\mu, \sigma)$ then $\log T \sim \mathcal{N}(\mu, \sigma)$.

Then we substitute the cumulative density function of the normal distribution into Equation (2.12), the details can be found in appendix A.2. The entries A_{11} and A_{22} are given by

$$A_{11} = \sum_{j=0}^{m+1} \frac{1}{P_j} \left(\frac{\partial P_j}{\partial \mu} \right)^2 - \frac{\partial^2 P_j}{\partial \mu^2}, \quad (2.38)$$

$$A_{22} = \sum_{j=0}^{m+1} \frac{1}{P_j} \left(\frac{\partial P_j}{\partial \sigma} \right)^2 - \frac{\partial^2 P_j}{\partial \sigma^2}. \quad (2.39)$$

For this data example, the maximum likelihood estimates are $\hat{\mu} = 0.1178$ and $\hat{\sigma} = 0.2886$, the 95% confidence interval for parameters μ and σ are $[-0.0947, 0.3303]$ and $[0.2003, 0.3769]$ respectively.

2.3.3.1.2 Reliability analysis

For Log-normal distribution, the reliability and the failure rate are given respectively by

$$R(t) = \int_{\log(t)}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\left(\frac{x-\mu}{\sqrt{2}\sigma}\right)^2} dx, \quad (2.40)$$

$$\lambda(t) = \frac{f(t)}{R(t)}. \quad (2.41)$$

With the distribution estimated above, the MTTF is equal to 117.29%. The failure rate behaviour obtained with Log-normal distribution is different from that of the Weibull distribution, the risk of failure quickly increases to a maximum then decreases (see Figure 2.7).

2.3.3.2 Non-homogeneous grouped data

The MLE is also done with the non-homogeneous grouped data presented in appendix B. The estimates are $\hat{\mu} = 0.6170$ and $\hat{\sigma} = 1.2965$.

2.3.3.2.1 Fisher information and confidence interval

For non-homogeneous grouped data, the Fisher information is given by Equation (2.32). With this data example, the 95% confidence intervals are $[-0.7839, 2.0180]$ for μ and $[0.6522, 1.9408]$ for σ .

2.3.3.2.2 Reliability analysis

The reliability function $R(t)$ and the failure rate $\lambda(t)$ are presented and compared with Weibull distribution (Figure 2.7) in section 2.3.4.2. It is observed that the change of parameter σ doesn't influence much the shape of failure rate $\lambda(t)$.

2.3.4 Comparison of estimations obtained by Weibull and Log-normal distributions

2.3.4.1 Homogeneous grouped data

Here probability densities estimated for Weibull and Log-normal distributions are compared in the Figure 2.4 with the data presented in Figure 2.3. From this figure, the Log-normal distribution gives a more skewed estimation for a normalized time closed to 100%. For the Weibull distribution, the estimation leads to a more symmetric PDF. With Log-normal distribution a shorter fatigue lifetime is estimated compared to Weibull distribution. Regarding the variance calculated for those two distributions, the variance value of Weibull distribution is smaller than that of Log-normal distribution, which means that the Weibull distribution is a better model for those data.

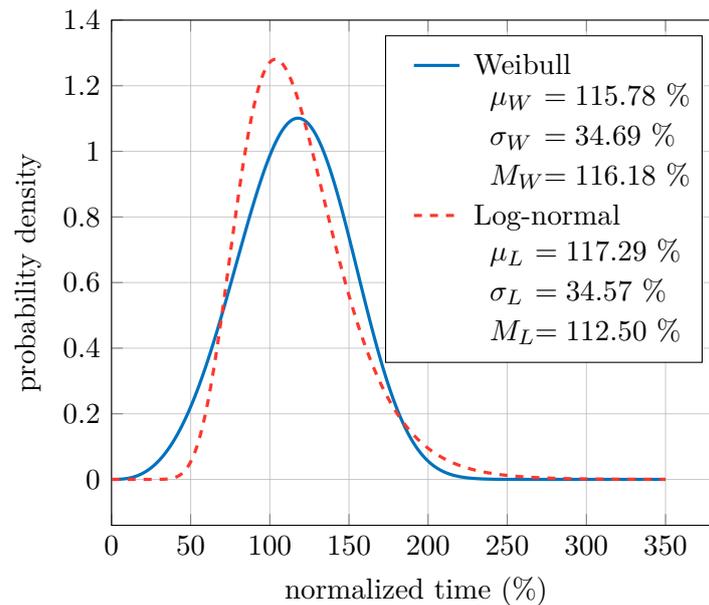


FIGURE 2.4 – PDF of the Weibull ($\hat{\alpha} = 1.2832, \hat{\beta} = 3.6875$) and Log-normal ($\hat{\mu} = 0.1178, \hat{\sigma} = 0.2886$) distributions

The reliability function $R(t)$ and failure rate $\lambda(t)$ of those two distributions are also calculated and presented in Figure 2.5. Up to 100%, the Log-normal distribution gives a better reliability. It is observed that for the same data, the Log-normal and Weibull distribution give similar results for the reliability function, but the failure rate functions are very different.

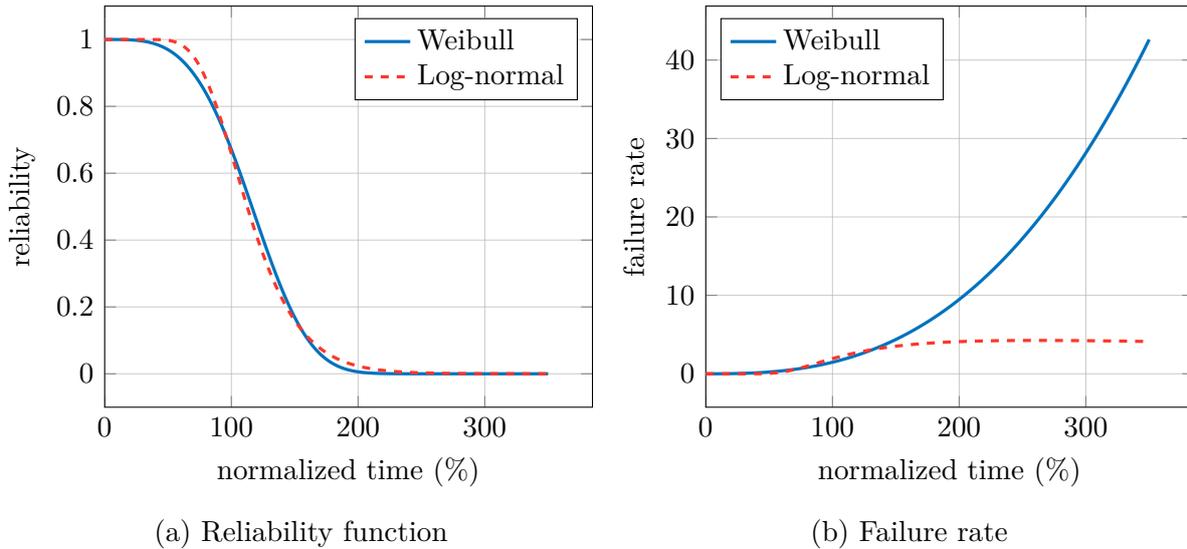


FIGURE 2.5 – Reliability analysis with Log-normal and Weibull distributions for homogeneous grouped data

2.3.4.2 Non-homogeneous grouped data

For non-homogeneous grouped data, a comparison of estimated Weibull and Log-normal distributions is also done. In Figure 2.6, the PDF are drawn. The medians for the two distributions are very close. The mean and standard deviation for Log-normal distribution are higher than those for the Weibull distribution, normally it means that a significant variability is observed with Log-normal distribution compared to Weibull distribution. But from the PDF of those two distributions (Figure 2.6), it seems that the Log-normal distribution has less variability than the Weibull distribution. This phenomenon is caused by the heavy-tailed distribution, which is the probability distribution whose tails are not exponentially bounded. It means that a density function goes to zero less rapidly than an exponential function [54]. It represents the case where the large value of a *r.v.* takes an important part. The Log-normal and Weibull distributions are both heavy-tailed distribution. Here in the Figure 2.6, the probability density of Log-normal distribution has a more centred density for smaller value of time, but it has a heavier right tail than Weibull distribution, which causes its larger value of mean and standard deviation. It also shows that the first or second moment (expectation and variance) are not a reliable indicators to compare those two distributions, the median can be a better choice.

The reliability and failure rate of the two distributions are also compared in Figure 2.7. The Log-normal model results in a lower degradation after 200%. For Log-normal distribution, the failure rate often increases at the beginning of product life and decreases after. For Weibull distribution, the forms of failure rate are very different depending on the shape parameter β .

In fact, only one distribution with one failure mode is usually not sufficient in order to model the failure rate of a complex structure. For a system, the bathtub-shaped failure rate function [55] is composed by three stages. The early life failure is often caused by defects in materials or manufacturing defects, then a flat region represents the stable function stage

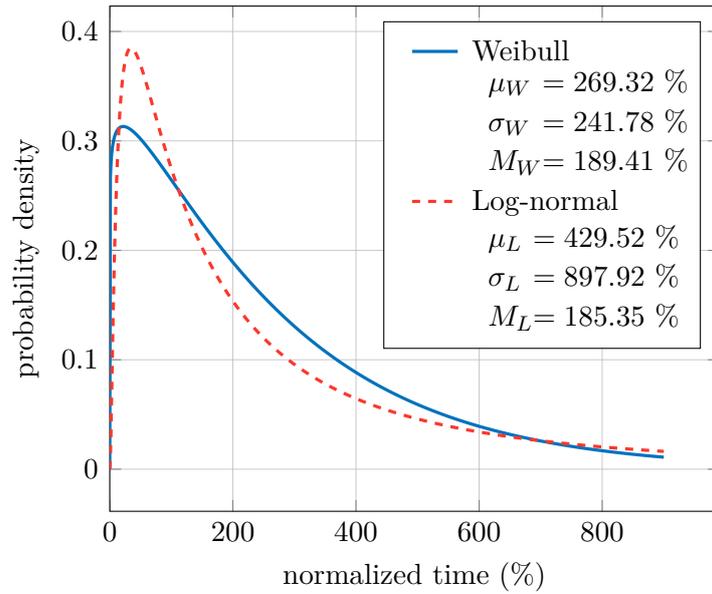


FIGURE 2.6 – PDF of Weibull ($\hat{\alpha} = 2.6651, \hat{\beta} = 1.0733$) and Log-normal distributions ($\hat{\mu} = 0.6171, \hat{\sigma} = 1.2965$)

of the system. The last part of increasing failure rate is caused by cumulative damage such as material fatigue. Thus it is a system with several failure modes and sometimes requires the mixed distribution to model the complex form. For example, a Log-normal modified Weibull distribution is proposed by Mohammed et al. [56]. For a mechanical structure, it is not realistic that the failure rate decreases with the time after the early stage. At the middle and ending phase of fatigue lifetime, the failure rate should normally increase. For this reason, some engineers do not use the Log-normal distribution to model the fatigue lifetime.

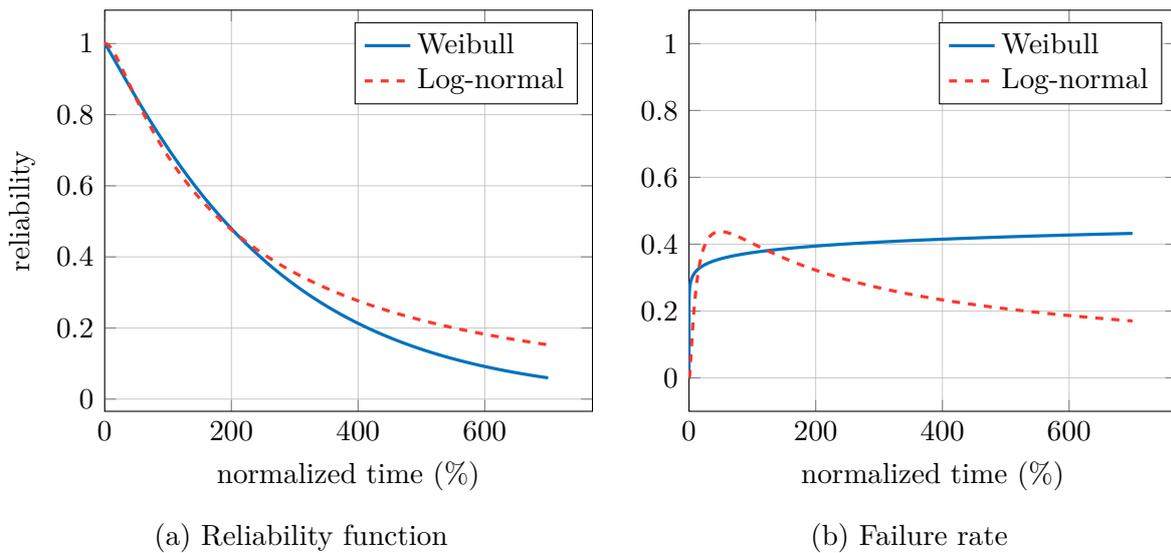


FIGURE 2.7 – Reliability analysis with Log-normal and Weibull distributions for non-homogeneous data

2.4 Goodness of fit tests for grouped data

Usually, the Weibull and Log-normal distributions are used to describe the scatter of mechanical structure fatigue lifetime. It is now expected to know which distribution fits better those data, thus the goodness of fit test is investigated.

The goodness of fit test is a type of statistical test to determine whether a sample follows a given distribution. Most widely known tests are χ^2 test, Kolmogorov-Smirnov test (KS) and Cramér-von Mises test (CVM). The null hypothesis H_0 and the alternative hypothesis H_1 are usually as follows :

H_0 : The sample follows the distribution F_0

H_1 : The sample does not follow the distribution F_0

Test statistics are calculated from the observed data for different statistical tests, and compared with a critical value (significant level) α . For a confidence level $1 - \alpha$, if the calculated statistic is greater than the critical value, the null hypothesis is rejected on this significant level. By definition, the critical level is the probability of getting a value of the test statistic as extreme as the observed value when H_0 is true. In other words, α is the smallest level of significance at which H_0 may be rejected for the observed sample. The critical values can be found in statistics table, they are percentiles of the asymptotic distribution of test statistic.

The χ^2 test for grouped data can be convenient, because it requires to divide the data in several intervals and leads to a discrete description. As for KS test, the commonly used statistics table is no longer adapted when defining the critical value for grouped data. The approximate asymptotic distribution of test statistic for grouped data or discontinuous distribution is not the same as the continuous data. Conover [57] proposed a method to calculate critical levels corresponding to selected critical values for discontinuous distribution, a statistic table of KS test for grouped data with N sample size and k intervals (see [58]). An application of Conover's procedure on KS test for grouped data is presented by Darmosiswoyo [59]. Besides, another approach by Smirnov transformation and randomization to transform discrete or grouped data to continuous uniform distribution is introduced [60]. Conover's procedure is used to do the hypotheses test for grouped data in the following sections.

2.4.1 Kolmogorov-Smirnov test for grouped data

The Conover's procedure is here applied on experimental data presented in Figure 2.3.

2.4.1.1 Kolmogorov-Smirnov test for continuous data

In general, for a random variable X , the Kolmogorov-Smirnov statistic is the greatest difference between empirical cumulative distribution F_n and the theoretical cumulative distribution F_0 with the selected group intervals. For one-sided Kolmogorov-Smirnov, the statistic is

$$D^- = \max(F_0(x) - F_n(x)) \quad (2.42)$$

or

$$D^+ = \max(F_n(x) - F_0(x)) \quad (2.43)$$

The two-sided Kolmogorov Smirnov distribution statistic is

$$D = \sup_{x \in \mathbb{R}} |F_n(x) - F_0(x)| \quad (2.44)$$

If the theoretical distribution is discrete, or the data are discrete or grouped, the asymptotic distribution of KS statistic is changed, and the statistics table is no longer usable. Conover proposed a method to treat grouped data [57, 59].

2.4.1.2 Conover's procedure

Suppose there are n samples, they are grouped in m intervals, the realisations of $r.v.$ X is arranged in increasing order, the probability of each interval can be calculated, thus the empirical cumulative distribution of grouped data F_n (blue full line in Figure 2.8) is obtained. The continuous theoretical distribution is transformed to discrete distribution F_0 (red dashed line in Figure 2.8) with m intervals, for each interval the probability is calculated by the integral of probability density in this interval.

The statistics D^- , D^+ and D are computed, then the critical level is computed. D^- is taken as an example to illustrate the calculation of critical level. The value of statistic D^- is d^- .

Firstly, the probability c_j for $0 \leq j < n$ corresponding to n with increasing ordered realisations of X is calculated as follows. Draw a horizontal line l_j with the ordinate $d^- + j/n$ as in Figure 2.8. If l_j does not intersect F_0 at a jump, $c_j = 1 - (d^- + j/n)$. If the line l_j intersects F_0 at a jump, then c_j equals to 1 minus the height of the top of the jump of F_0 .

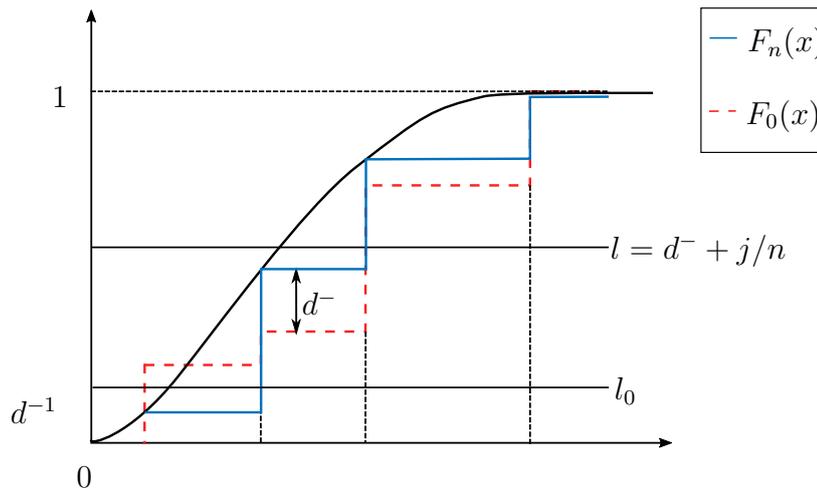


FIGURE 2.8 – Schematic of the KS test for the grouped data

Secondly, the constants b_j are calculated with $b_0 = 1$ and $b_i = 1 - \sum_{i=0}^{j-1} \binom{j}{i} c_j^{j-i} b_i$, where $j \geq 1$.

Lastly, the critical level of d^- is computed by

$$P(D^- \geq d^-) = \sum_{j=0}^n \binom{n}{j} c_j^{n-j} b_j. \quad (2.45)$$

The computation of $P(D^+ \geq d^-)$ is similar. The horizontal line l_j is drawn with ordinate $1 - d^+ - j/n$. If l_j doesn't intersect F_0 at a jump, f_j is equal to $1 - d^+ - j/n$. If the line l_j intersects F_0 at a jump, then f_j is equal to 1 minus the height of the bottom of the jump of F_0 . The constants e_j are respectively equal to $e_0 = 1$ and $e_i = 1 - \sum_{i=0}^{j-1} \binom{j}{i} f_i^{j-i} e_i$ with $j \geq 1$. Finally the critical level is calculated by

$$P(D^+ \geq d^+) = \sum_{j=0}^n \binom{n}{j} f_j^{n-j} e_j. \quad (2.46)$$

As for two-sided test, the probability $P(D \geq d)$ is approximated by Equation (2.47), with $d = \max(d^-, d^+)$ [57, 59]

$$\begin{aligned} P(D \geq d) &= P(D^- \geq d \cup D^+ \geq d) \\ &= P(D^- \geq d) + P(D^+ \geq d) - P(D^- \geq d \cap D^+ \geq d), \end{aligned} \quad (2.47)$$

where

$$P(D^- \geq d \cap D^+ \geq d) = P(D^- \geq d)P(D^+ \geq d | D^- \geq d) \leq P(D^- \geq d)P(D^+ \geq d).$$

If $D^- \geq d$ and $D^+ \geq d$ are independent,

$$P(D^- \geq d \cap D^+ \geq d) = P(D^- \geq d)P(D^+ \geq d).$$

In addition, if F_0 is a symmetric distribution, approximately we have

$$P(D^- \geq d^-) = P(D^+ \geq d^+),$$

thus

$$P(D \geq d) = 2P(D^- \geq d^-) - P^2(D^- \geq d^-)$$

or

$$P(D \geq d) = 2P(D^+ \geq d^+) - P^2(D^+ \geq d^+).$$

If the critical level is smaller than the significant level of the test, the null hypotheses H_0 is rejected.

2.4.1.3 Result for simulated data of Log-normal distribution

The test is firstly applied on simulated data. The simulated uncensored data sample for a size $n = 50$ are generated with Log-normal distribution, $\mu = 0.1$ and $\sigma = 0.2$. Then the data are grouped in 20 intervals, the grouped data sample is described as follows :

$$\begin{aligned} t &= [0; 0.1; 0.2; 0.3; 0.4; 0.5; 0.6; 0.7; 0.8; 0.9; 1.0; 1.1; 1.2; 1.3; 1.4; 1.5; 1.6; 1.7; 1.8; 1.9; 2.0] \\ n &= [0; 0; 0; 0; 0; 0; 1; 3; 6; 5; 10; 9; 6; 4; 3; 1; 1; 1; 0; 0] \end{aligned}$$

where t is the time interval and n is the number of failures in each time interval.

The theoretical and empirical cumulative distribution are presented in Figure 2.9. The KS statistic equals to 0.0379, as the Log-normal distribution is not symmetric, the $P(D^- \geq d)$ and $P(D^+ \geq d)$ should be calculated separately.

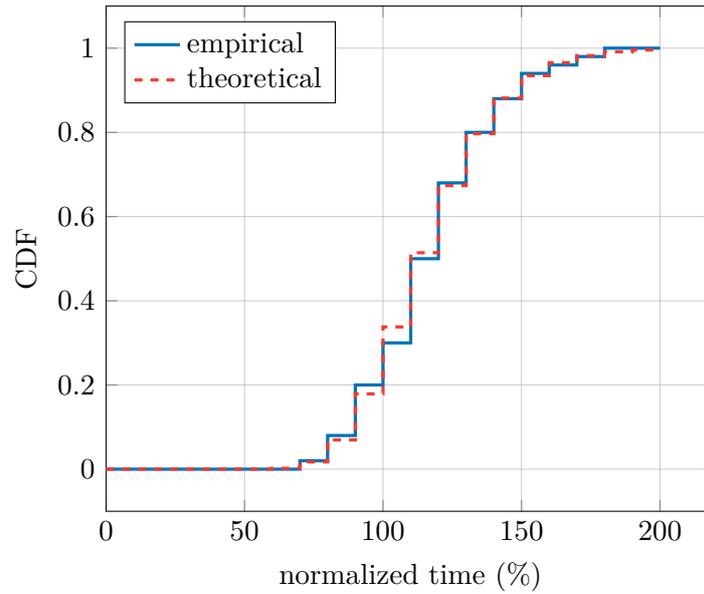


FIGURE 2.9 – Theoretical and empirical cumulative distributions of simulated grouped data sample from Log-normal distribution

According to Conover’s procedure the computation has been done

$$P(D^- \geq d) = 0.6655$$

and

$$P(D^- \geq d) = 0.6875,$$

thus the critical value is

$$P(D \geq d) = P(D^- \geq d) + P(D^+ \geq d) - P(D^- \geq d)P(D^+ \geq d) = 0.4379.$$

The critical level is greater than 0.4, thus the null hypothesis H_0 is accepted on 0.40 significance level. Consequently, the data follow the Log-normal distribution.

2.4.1.4 Result for industrial experimental data of Log-normal distribution

The test is applied on Renault experimental data presented in Figure 2.3. The MLE of grouped data is done for Log-normal and Weibull distributions, then the test is applied for those two distributions with the parameters estimated.

For Log-normal distribution, the comparison of the empirical distribution and the estimated theoretical CDF in form of grouped data is drawn in Figure 2.10a. The statistic equals to 0.1584, then the critical level is computed

$$P(D^- \geq d) = 0.0920$$

and

$$P(D^- \geq d) = 0.0904,$$

thus the critical value is

$$P(D \geq d) = P(D^- \geq d) + P(D^+ \geq d) = 0.1824.$$

The critical level is greater than 0.15, thus we can say the null hypothesis H_0 (the data follow the Log-normal distribution) is accepted on 0.15 significance level.

For Weibull distribution, the comparison of the empirical distribution and the estimated theoretical CDF in form of grouped data is drawn in Figure 2.10b. The greatest distance between the empirical cumulative distribution and the theoretical distribution is equal to 0.1711. The one-sided critical level is $P(D^- \geq d) = 0.0506$ and $P(D^+ \geq d) = 0.0679$, thus the critical value is $P(D \geq d) = P(D^- \geq d) + P(D^+ \geq d) = 0.1185$. The critical level is greater than 0.1, thus the null hypothesis H_0 (the data follow the Weibull distribution) is accepted on 0.10 significance level.

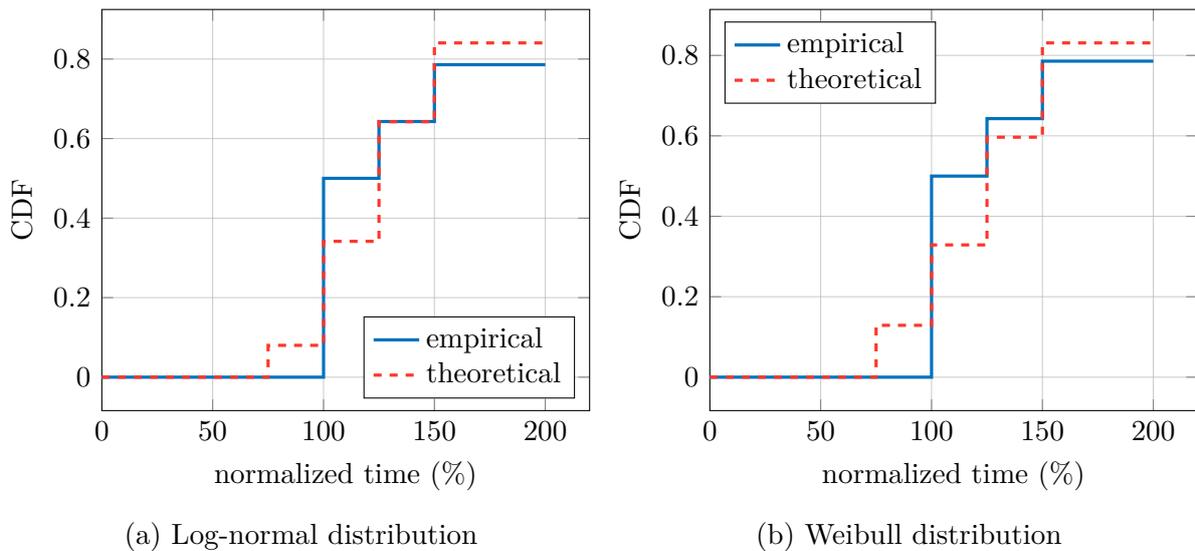


FIGURE 2.10 – Theoretical and empirical cumulative distributions of experimental grouped data for Log-normal and Weibull distributions

Comparing the test results for Log-normal and Weibull distributions, the Log-normal distribution fits better this set of data. According to the expert experience, the Log-normal distribution is chosen to describe the fatigue lifetime of the studied structure, the test results are in accordance with this practice.

2.5 Bayesian estimation for Weibull distribution and Log-normal distribution

The Bayesian approach takes into account both the information obtained from the observations and the prior information, which is also the main difference between frequentist (MLE) and Bayesian method. According to [61], Bayes theorem weights the prior information and data according to their relative strengths in the posterior distribution. With the increase of observations, the influence of prior information becomes smaller and smaller. So with the decrease of observations, the Bayesian estimation can lead to a result less influenced by the measurement error.

2.5.1 Method of Bayesian estimation

Let D represent the observed data and θ represent the model parameters, θ is a random variable which follows a prior distribution $\pi_{pr}(\theta)$, the posterior distribution of θ given the data is

$$p(\theta|D) = \frac{L(D|\theta)\pi_{pr}(\theta)}{p(D)} = \frac{L(D|\theta)\pi_{pr}(\theta)}{\int L(D|\theta)\pi_{pr}(\theta) d\theta}, \quad (2.48)$$

where $L(D|\theta)$ is the likelihood function and $p(D)$ the marginal distribution of D .

Usually there are two types of prior information, subjective and objective [62]. For example, the opinion of experts in a domain is a subjective prior information, the uniform distribution and Jeffreys prior are the objective prior commonly used. The proper choice of the prior information is a controversial topic in Bayes theorem.

During the calculation of Equation (2.48), usually the denominator (marginal distribution) $p(D)$ is an integral difficult to calculate analytically. Some researchers work on approximation methods [63] [45]. Some numerical methods are popular to evaluate the posterior distribution, such as the classical Monte Carlo method (importance sampling), and the Markov chain Monte Carlo method [64] [65]. A survey of the major techniques and available approaches for the approximation of integrals is proposed by Evans et al. [66] : techniques in asymptotic approximation as Laplace's method, the importance sampling, adaptive importance sampling, multiple quadrature and Markov chain methods; these methods are applied on three examples. The Markov chain method is widely used due to the simplicity of construction and rapid convergence in some cases.

The importance sampling and Markov chain Monte Carlo methods are presented as follows.

2.5.1.1 Monte Carlo method (importance sampling)

To calculate the integral $I = \int h(x)dx$ with Monte Carlo method, the integrated part is transformed as follows

$$I = \int \frac{h(x)}{f(x)} f(x) dx, \quad (2.49)$$

where $f(x)$ is a PDF and its support is included in those of h . The above form corresponds to the expectation of $\frac{h(X)}{f(X)}$. According to the law of large numbers

$$P \left(\lim_{n \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \frac{h(X_i)}{f(X_i)} = I \right) = 1, \quad (2.50)$$

the expectation could be estimated numerically as the mean value when the sample size n is large enough. It is important to choose a proper $f(x)$, for example the same as $h(x)$ in order to obtain an accurate approximation result.

2.5.1.2 Markov chain Monte Carlo

The Markov chain Monte Carlo (MCMC) method is an application of Markov chain in general state space [65]. The construction of Markov chains can be on discrete space [67] or on general (non-countable) state space. A random sequence is a Markov chain if

its future value knowing the past and the present depends only on the present. A Markov chain is characterised by the transition kernel, which is the transition probability from one state to another. The Markov chain has some properties such as the convergence, the limiting law, memorylessness of pass and the initial law defining the distribution of the first state of a chain.

A probability density π_d is defined on a general state space (Ω, \mathcal{X}) . A transition on (Ω, \mathcal{X}) is an application $P : (\Omega, \mathcal{X}) \rightarrow [0, 1]$, $P(\cdot, A)$ for all $A \in \mathcal{X}$ is measurable and $P(x, \cdot)$ is a probability for all $x \in \Omega$. A probability measure $\pi(\cdot)$ on \mathcal{X} is given by [65] and expressed as

$$\pi(A) = \frac{\int_A \pi_d(x) dx}{\int_{\mathcal{X}} \pi_d(x) dx}. \quad (2.51)$$

The MCMC method uses the stationary distribution property of a Markov chain. It is expected to simulate a Markov chain with a stationary distribution π . A reversible Markov chain is stationary and an ergodic Markov chain has a unique stationary distribution same as its limiting distribution. Thus to apply the MCMC, it is necessary that the Markov chain in general space is reversible and ergodic. A chain is ergodic if it is ϕ -irreducible (a weaker condition than for finite \mathcal{X}), aperiodic and Harris recurrent. ϕ -irreducible means any state of a chain is reachable from anywhere; the aperiodicity means that the chain will not oscillate between disjoint subsets $\mathcal{X}_1, \mathcal{X}_2, \dots, \mathcal{X}_d \subseteq \mathcal{X}$ with a period $d \geq 2$. Harris recurrent means that for all subset $B \subseteq \mathcal{X}$ with $\pi(B) > 0$ and all $x \in \mathcal{X}$, the chain will eventually reach B from x with probability 1. The latter condition is stronger than ϕ -recurrent, because it requires also the transition kernels $P(x, \cdot)$ are absolutely continuous with respect to $\pi(\cdot)$ [65].

To formulate the MCMC method, the Markov chain has the transition probabilities $P(x, B)$, $x \in \Omega$ from x to a subset B . Its stationary distribution $\pi(\cdot)$ yields the following equation

$$\int_x \pi(dx) P(x, B) = \pi(B). \quad (2.52)$$

A Markov chain is reversible if the transition kernels submits to the detailed balance condition

$$\pi(dx) P(x, dy) = \pi(dy) P(y, dx). \quad (2.53)$$

If Markov chain is reversible with respect to $\pi(\cdot)$, then $\pi(\cdot)$ is the stationary distribution of the chain with

$$\int_{x \in \mathcal{X}} \pi(dx) P(x, dy) = \int_{x \in \mathcal{X}} \pi(dy) P(y, dx) = \pi(dy). \quad (2.54)$$

In the background of Bayesian method, if the posterior distribution is considered as the stationary distribution, we can generate the sample following the posterior distribution, then calculate the estimators from the sample.

The sample points generated of X converge to the stationary distribution after some iterations of sampling, the sample discarded before convergence is called the burn-in [68].

According to [69], it is suggested that the initial 1% or 2% of runs is usually sufficient for a large n of total sample.

The most common algorithms to generate Markov chain Monte Carlo are the Metropolis-Hastings sampling and Gibbs sampling.

2.5.1.2.1 Method of Metropolis-Hastings (MH)

The Metropolis-Hastings sampler obtains next state dy of the chain from state x by sampling a candidate point from a proposal distribution $Q(x, \cdot)$, whose PDF is $q(x, \cdot)$. The Markov chain should be ergodic, usually the condition is met when the proposal distribution has a positive density on the same support as the target distribution [68].

Suppose ζ is the acceptance rate of a sampling, the detailed balance condition in Equation (2.53) can be rewritten as

$$\pi(dx)q(x, dy)\zeta(y, dx) = \pi(dy)q(y, dx)\zeta(x, dy). \quad (2.55)$$

In the case where the acceptance rate is too small, then the convergence of Markov chain could be very slow. So we amplify $\zeta(y, dx)$ and $\zeta(x, dy)$ at the same time until the larger one equals to 1, thus the acceptance rate could be written in the following form

$$\zeta(y, dx) = \min \left(1, \frac{\pi(dy)q(y, dx)}{\pi(dx)q(x, dy)} \right). \quad (2.56)$$

The choice of PDF q is important for the convergence rate. If the step is too small, it takes too much time to travel all the support of target distribution. If the step is too large, the acceptance rate can be too low. So a proposal distribution q should have a proper scale compared to target distribution.

2.5.1.2.2 Random Walk Metropolis-Hastings

The random walk Metropolis-Hastings is the original version of the MH algorithm. This version is introduced in the book of Art B. Owen [70], which means the transition law is taken as symmetric distribution. The transition distribution proposes a jump V with $V := Y - X$ to obtain the new value, and $q(v) = q(-v)$ for random walk Metropolis-Hastings [71]. For example, if the normal distribution $Q(x, \cdot) = N(x, \sigma^2)$ is chosen, then $q(y, dx) = q(x, dy)$ and the acceptance rate becomes

$$\zeta(x, dy) = \min \left(1, \frac{\pi(dy)}{\pi(dx)} \right). \quad (2.57)$$

To avoid taking an initial value far from the real estimates, the initial x_0 is taken as the maximum likelihood estimate.

Now, the generation algorithm of a *r.v.* X following the stationary distribution of a Markov chain is presented.

Generation algorithm

1. Initialise $X = x_0$ (take the MLE estimator for example)
2. Generate y from x_i with the proposed distribution $q(x, dy)$
3. Calculate the acceptance rate $\zeta(x, dy)$ with Equation (2.56)
4. Generate a *r.v.* $U \sim Unif[0, 1]$,
if the value $u > \zeta(x, dy)$
accept y and $x_{i+1} = y$
else $x_{i+1} = x_i$
5. Repeat Step2 and Step3 n times
6. Take the samples from $\frac{n}{10}$ to n times to avoid 'burning in' section
7. End

2.5.1.2.3 Gibbs sampling

Gibbs sampling is a special case of Metropolis-Hastings algorithm with the acceptance rate equal to 1. Suppose two Markov chain Z_1 and Z_2 to be sampled, the generation algorithm is presented next.

Generation algorithm

1. Initialize $Z_1 = z_{1,0}$ and $Z_2 = z_{2,0}$
2. Generate $Z_1 = z_{1,i}$ from $P(Z_1|Z_2 = z_{2,i-1})$
3. Generate $Z_2 = z_{2,i}$ from $P(Z_2|Z_1 = z_{1,i})$
4. $i=i+1$ and repeat steps 2 through 4 n times
5. Take the samples from $\frac{n}{10}$ to n times to avoid 'burning in' section
6. End

The Gibbs sampler is easier to calculate for multi-dimensional sampling, because it allows to generate each parameter separately but not generate all the parameters on time and decide whether accept them or not as the Metropolis-Hastings method. However, it requires the full conditional distribution of one random variable with respect to the others, which is not always available.

2.5.2 Bayesian estimation for Weibull distribution

The conjugate prior is usually considered as the priority when performing the Bayesian estimation. However, it is known that the natural conjugate prior of two parameters α and β in Weibull distribution doesn't exist. But when the shape parameter β is known, the conjugate prior of α is inverse gamma distribution [72]. Researchers choose different prior distribution in their work. In the case of [73], the prior information is taken in agreement with the experts in industry. In the paper [46], the prior distributions of parameter α and β are both taken as the gamma law; at first the hyper-parameters are equal to zero to make non-informative prior, and then the hyperparameters are given values, but finally the results are not very different in those 2 cases.

2.5.2.1 Bayesian estimation of one parameter

According to [45], γ is considered as a random variable and β is a fixed value.

2.5.2.1.1 Prior information

The prior distribution is considered to be inverse-gamma distribution for γ ,

$$\pi_{pr}(\gamma) = \frac{b^a}{\Gamma(a)} \gamma^{-(a+1)} e^{-\frac{b}{\gamma}}, \quad a > 0, b > 0, \quad (2.58)$$

where a and b are two parameters of inverse-gamma distribution.

To determine parameters a and b , we let the expectation of prior γ equal to $\hat{\gamma}$ (estimate of γ by MLE), the variance equal to a constant C , that is

$$\mathbb{E}(\gamma) = \frac{b}{a-1} = \hat{\gamma}_{mle}, \quad a > 1, b > 0, \quad (2.59)$$

$$Var(\gamma) = b^2 \left(\frac{1}{(a-1)^2(a-2)} \right) = C, \quad a > 2, b > 0. \quad (2.60)$$

With the prior information above, the posterior distribution is

$$p(\gamma | \underline{N}^f) = \frac{p(\underline{N}^f | \gamma) \pi_{pr}(\gamma)}{p(\underline{N}^f)} = \frac{\gamma^{-(a+1)} e^{-\frac{S}{\gamma}} \prod_{j=1}^k \left(1 - e^{-\frac{(t_j^\beta - t_{j-1}^\beta)}{\gamma}} \right)^{N_j^f}}{\int_0^\infty \gamma^{-(a+1)} e^{-\frac{S}{\gamma}} \prod_{j=1}^k \left(1 - e^{-\frac{(t_j^\beta - t_{j-1}^\beta)}{\gamma}} \right)^{N_j^f} d\gamma}. \quad (2.61)$$

Under the squared loss function, the estimator of γ is the expectation of posterior distribution

$$\hat{\gamma} = \mathbb{E}[\gamma | x] = \int \gamma p(\gamma | \underline{N}^f) d\gamma. \quad (2.62)$$

2.5.2.1.2 Calculation of posterior estimator by approximation approach and importance sampling

The experimental data have been presented in Figure 2.3. To perform the Bayesian estimation mentioned above, we proceed as follows.

At first the MLE is done, and the maximum likelihood estimators are obtained : $\hat{\alpha} = 1.2832$, $\hat{\beta} = 3.6875$, $\hat{\gamma} = 2.5080$. The prior distribution of γ is inverse-gamma distribution $\pi(\gamma) = \frac{b^a}{\Gamma(a)} \gamma^{-(a+1)} e^{-\frac{b}{\gamma}}$, with $a > 0$ and $b > 0$. To determine a and b , we assume that the prior expectation satisfies $\mathbb{E}_\pi(\gamma) = \hat{\gamma} = 0.6404$, and the prior variance is chosen as $V_\pi(\gamma) = 10$. Thus $a = 2.3144$ and $b = 3.2963$.

Two methods are used to calculate the posterior distribution and the estimator of γ under squared error function : the approximation method proposed in [45] and the Monte Carlo method. The estimators calculated by the two methods are presented in Table 2.1. The estimators of parameter γ obtained by these two methods are similar.

With importance sampling method, the Bayesian estimator of γ is equal to $\tilde{\gamma} = 2.4994$, the maximum likelihood estimator of β is adopted $\hat{\beta} = 3.6875$, then the Bayesian

TABLEAU 2.1 – Estimator $\tilde{\gamma}$ calculated by the approximation and importance sampling methods

	Approximation	Importance sampling
$\tilde{\gamma}$	2.0593	2.4994

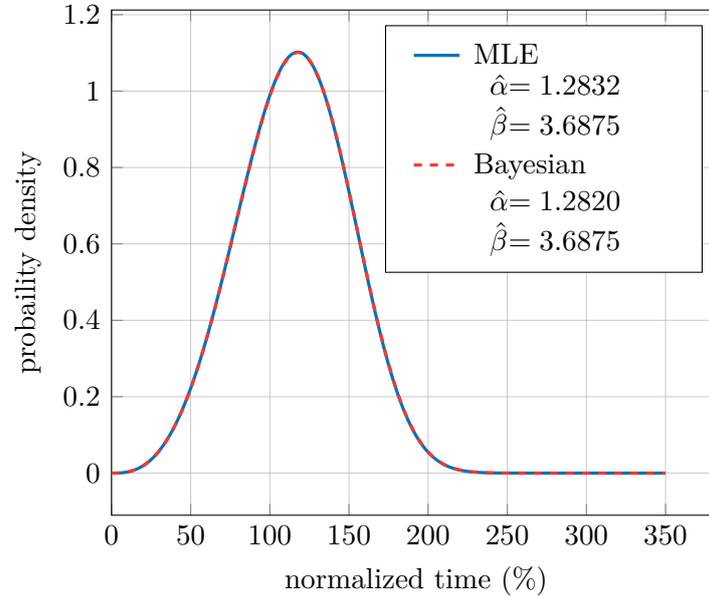


FIGURE 2.11 – PDF estimated by MLE and Bayes methods for Log-normal distribution

estimator of α is calculated by $\tilde{\alpha} = \tilde{\gamma}^{1/\hat{\beta}} = 1.2820$. The estimated probability density function by Bayesian and maximum likelihood are both presented in Figure 2.11.

The difference between the PDF estimated by Bayesian and MLE methods is given by the prior law of parameter. In general, when there are more data, the estimator is less influenced by the prior law but more influenced by observations. On the contrary, the prior law has more impact than observations. So in the case where there are not many data, the Bayesian estimation is a better choice to avoid the 'bad' influence of 'bad' observations.

2.5.2.2 Bayesian estimation of two parameters

2.5.2.2.1 Prior information

According to the expert opinions in companies, the prior information was considered to be the uniform distribution on the confidence interval calculated with MLE. With the observations of number of failures N^f shown in Figure 2.3, the Bayesian estimation is done.

The maximum likelihood estimates are $\hat{\alpha} = 1.2832$, $\hat{\beta} = 3.6875$, and their 95% confidence interval are $[1.0984, 1.4679]$ and $[2.0652, 5.3098]$ respectively. The prior distributions for parameters α and β are uniform distributions on their confidence interval separately, and they are independent. Denote the confidence intervals are $[a_1, b_1]$ and $[a_2, b_2]$ for α and β respectively, their prior distributions are

$$\pi_{pr}(\alpha) = \frac{1}{b_1 - a_1} \mathbb{1}_{[a_1, b_1]}(\alpha), \quad (2.63)$$

$$\pi_{pr}(\beta) = \frac{1}{b_2 - a_2} \mathbb{1}_{[a_2, b_2]}(\beta). \quad (2.64)$$

With the Bayes formula detailed in Equation (2.65), the posterior joint distribution of the parameters α and β is written as

$$p(\alpha, \beta | \underline{N}^f) = \frac{\pi_{pr}(\alpha) \pi_{pr}(\beta) \mathcal{L}(\underline{N}^f | \alpha, \beta)}{\int \int \pi_{pr}(\alpha) \pi_{pr}(\beta) \mathcal{L}(\underline{N}^f | \alpha, \beta) d\alpha d\beta}. \quad (2.65)$$

2.5.2.2.2 Calculation of posterior estimator by MCMC

The Metropolis-Hastings algorithm is used to sample the posterior distribution of α and β . The proposal distribution is chosen as normal distribution with a standard deviation equal to 0.1. The normalized histogram of posterior distribution sample of parameters α and β with MCMC is represented in Figure 2.12. The sample is generated with a size equal to 5000, the first 500 trials are neglected to ensure that only the converged part of the sample is taken into account.

The estimator can be taken as the mean value under squared error loss function or as the median under absolute value loss function. With the previous sample, the mean and median are obtained in table 2.2. The mean values are taken as the estimators for α and β .

TABLEAU 2.2 – Estimates of α and β under squared error loss function (mean) and absolute value loss function (median) with MCMC method

	mean	median
$\tilde{\alpha}$	1.2936	1.2906
$\tilde{\beta}$	3.6075	3.6001

The convergence of the MCMC algorithm depends on the initial value of parameters and the proposal distribution. In general, the trajectories of MCMC sample are drawn to verify the convergence [64]. In Figure 2.13, the trajectories of α and β posterior samples are presented. It is shown that the sample values are always around 1.3 for α and 3.5 for β , even at the beginning of the sampling. This is because the initial values take the MLE estimates and the prior distributions are uniform in the confidence interval of MLE estimator. Thus the prior information is extracted from the observed data with no more added information. The final Bayesian estimates of the parameters will not be very different from the MLE estimates. But, this kind of method to define the prior information is useful. When new tests observations are obtained, we can use the previous observations as prior information and the new information by the likelihood function to get a better estimation.

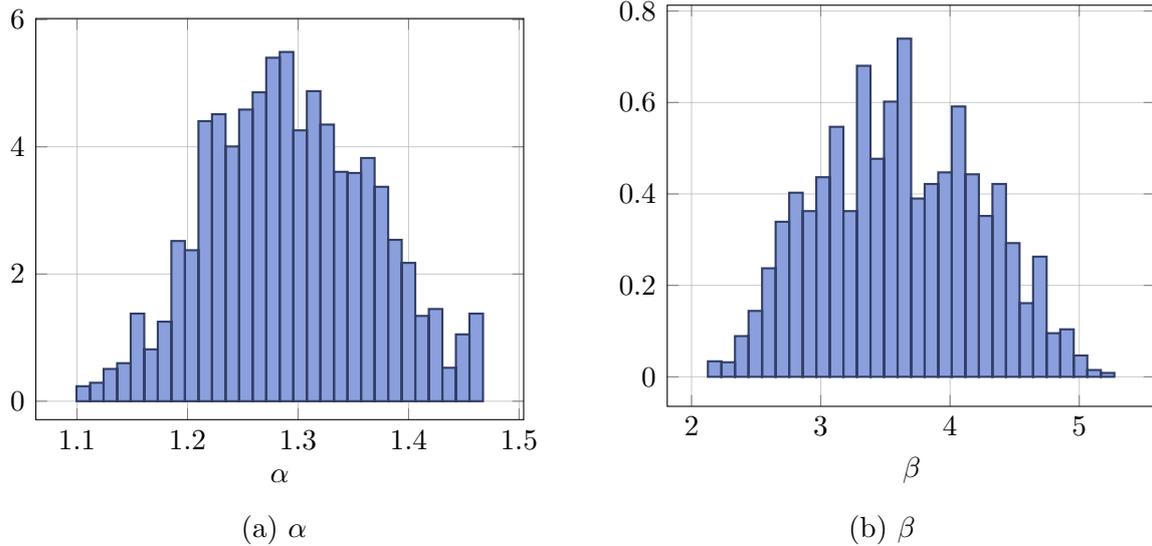


FIGURE 2.12 – Normalized histogram of posterior parameters

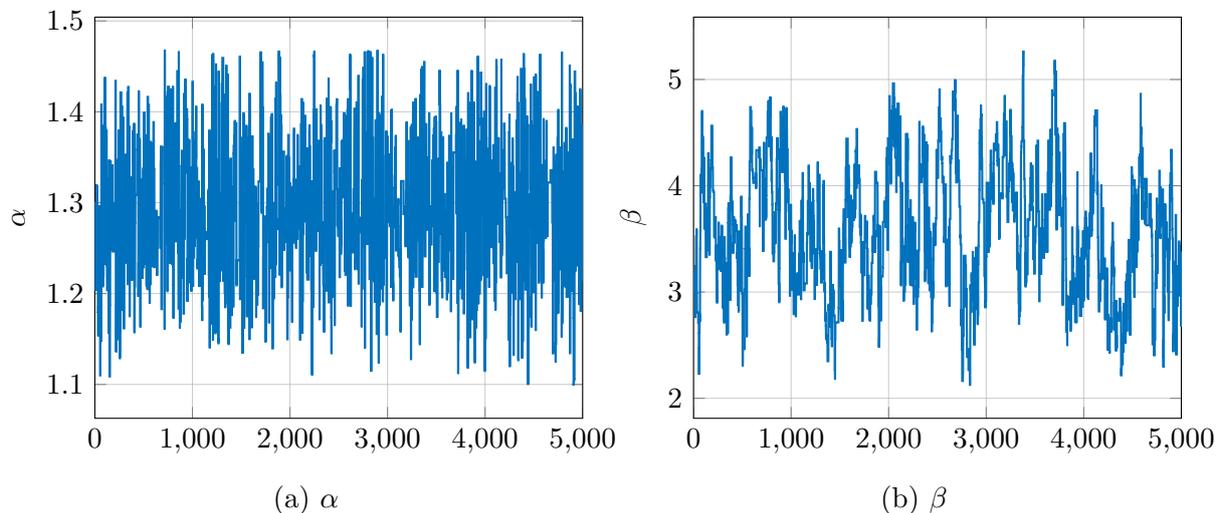


FIGURE 2.13 – Trajectory of MCMC sampling

2.5.3 Bayesian estimation for Log-normal distribution

The Log-normal distribution is also frequently used in the fatigue domain, and usually for the fatigue validation tests of automotive components, according to the experts' opinion of the Renault company.

The theory of Bayesian estimation is given in Section 2.5.1. In the case of Log-normal distribution, the cumulative density function is provided by Equation (2.34). The parameters to be estimated are μ and σ ; set $\theta = (\mu, \sigma)$.

2.5.3.1 Bayesian estimation of one parameter

At first, the Bayesian estimation is only done for the parameter μ , the maximum likelihood estimator $\hat{\sigma}$ is fixed and taken the value of maximum likelihood estimate.

2.5.3.1.1 Prior information

The normal distribution is chosen as the prior law of parameter μ , with the expectation equal to the maximum likelihood estimator. The variance is equal to 1, *i.e.*, $\pi_{pr}(\mu) \sim N(\hat{\mu}, 1)$.

2.5.3.1.2 Calculation of the posterior estimator by MCMC

The method of MH is used to generate the sample y following the posterior distribution $p(\mu|x)$,

$$p(\mu|x) = \frac{\mathcal{L}(x|\mu)\pi_{pr}(\mu)}{\int \mathcal{L}(x|\mu)\pi_{pr}(\mu) d\mu}, \quad \pi_{pr} = \mathcal{N}(\hat{\mu}, 1). \quad (2.66)$$

For the proposed distribution q , here we use the normal distribution with the expectation equal to the the last generated sample y_i and the standard deviation is chosen equal to 0.5 $q_i = \mathcal{N}(y_i, 0.5^2)$.

The Bayesian estimation is done for Log-normal distribution with the data presented in Figure 2.3. According to Section 2.3.3.1, the values of maximum likelihood estimators are $\hat{\mu} = 0.118$ and $\hat{\sigma} = 0.289$. As mentioned above, the prior information of the parameter μ is taken as normal distribution $\pi_{pr}(\mu) \sim \mathcal{N}(\hat{\mu}, 1)$, and the posterior distribution of μ is sampled by the algorithm Metropolis-Hastings with a sample size equal to 5000. The normalized histogram of the sample of μ posterior distribution is presented in Figure 2.14. The estimator is chosen as the mean value under squared error loss function, the estimator $\tilde{\mu}$ is equal to 0.119. And for the parameter σ , the maximum likelihood estimator $\hat{\sigma}$ is taken.

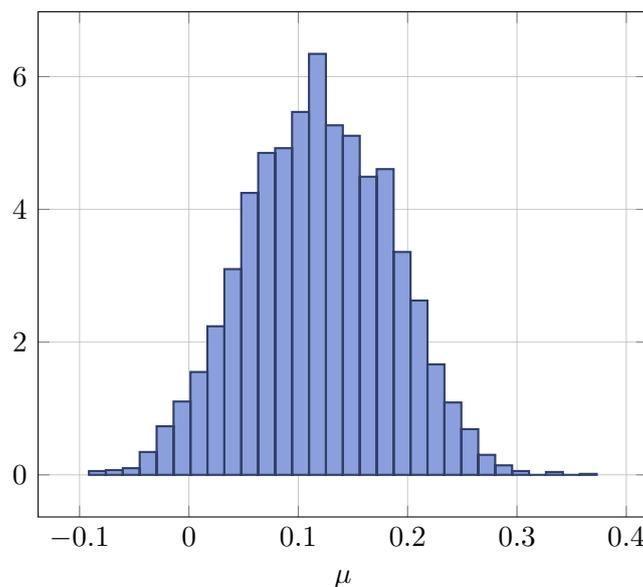


FIGURE 2.14 – Normalized histogram of posterior sample of μ by MCMC

2.5.3.1.3 Calculation of the integral in the Bayes formula by importance sampling

Importance sampling is a method used for calculating integral by the principle of Monte Carlo simulation. The posterior distribution $p(\mu|x)$ is given in Equation (2.66), the integral $I = \int \mathcal{L}(x|\mu)\pi_{pr}(\mu) d\mu$ corresponding to the denominator has to be calculated.

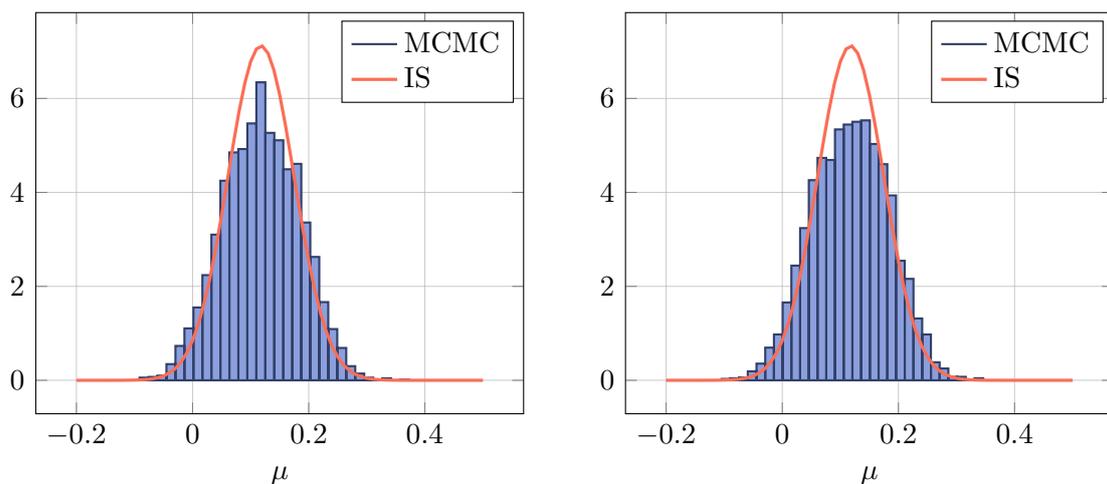
According to the strong law of large numbers, we have

$$\frac{1}{n} \sum_{i=1}^n \frac{\mathcal{L}(x|\mu)\pi_{pr}(\mu)}{g(\mu)} \xrightarrow[n \rightarrow +\infty]{a.s.} \int \mathcal{L}(x|\mu)\pi_{pr}(\mu) d\mu, \quad (2.67)$$

where $g(\mu)$ is the sampling distribution of μ and is chosen the same as $\pi_{pr}(\mu)$ for reasons of efficiency and simplified calculation technique.

2.5.3.1.4 Comparison of posterior distributions of parameter μ obtained by MCMC and importance sampling

Here we use two methods to determine posterior parameter distribution, the MH algorithm of MCMC and Monte Carlo simulation (importance sampling). Two different proposal distributions of MCMC are tested separately for $q_i = \mathcal{N}(y_i, 0.5^2)$ and $q_i = \mathcal{N}(y_i, 5^2)$. The posterior samples are presented in Figure 2.15 for these two values of standard deviation, and compared with the distribution obtained by importance sampling. It is observed that the sample obtained with a proposal distribution for a standard deviation equal to 0.5 is closer to the results obtained by importance sampling than that with a proposal distribution for standard deviation equal to 5, especially in the middle part of the distribution. It shows that a proper step length proposed by the proposal distribution plays important role to obtain satisfactory MCMC sample, it can also influence the convergence speed [74].



(a) Proposal distribution with standard deviation equal to 0.5

(b) Proposal distribution with standard deviation equal to 5

FIGURE 2.15 – Posterior parameter μ sampled by MH and calculated by importance sampling

A performance comparison between the MCMC and importance sampling can be found in [75]. It is showed for one example that importance sampling result is comparable to that of MCMC. It is concluded that the importance sampling can be used as an alternative method to the MCMC method.

2.5.3.2 Bayesian estimation of two parameters

2.5.3.2.1 Normal-inverse gamma prior

With a log transformation applied on the data, the estimation of Log-normal distribution could be considered as the estimation of normal distribution. The natural conjugate prior distribution of normal distribution is the Normal-inverse-gamma distribution [76], the proof of the conjugate prior distribution is presented in appendix C.

Suppose $\tau = \frac{1}{\sigma^2}$, the prior distribution for μ and τ is then Normal-Gamma distribution : $\tau \sim \Gamma(\alpha, \beta)$, with $\alpha > 0, \beta > 0$, and conditionally on τ , the parameter $\mu \sim \mathcal{N}(\nu, \frac{1}{k\tau})$, with $k > 0, \nu \in \mathbb{R}$

$$\begin{aligned} \pi_{pr}(\mu, \tau) &= \pi_{pr}(\mu|\tau)\pi_{pr}(\tau) \\ &= \frac{\beta^\alpha}{\Gamma(\alpha)} \tau^{\alpha-1} e^{-\beta\tau} \frac{1}{\sqrt{2\pi}} (k\tau)^{1/2} \exp\left[-\frac{k\tau}{2}(\mu - \nu)^2\right] \\ &\propto \tau^{\alpha-1/2} \exp\left[-\tau\left[\beta + \frac{k}{2}(\mu - \nu)^2\right]\right] \end{aligned} \quad (2.68)$$

The prior law Normal-Gamma is $\mathcal{N} - (\nu, \frac{1}{k\tau}, \alpha, \beta)$. Data of the Figure 2.3 are chosen, the maximum likelihood estimators are used to define the parameters in the prior distribution law, $\hat{\mu} = 0.1178, \hat{\sigma} = 0.2887$. Respectively, 95% confidence intervals are $[-0.0947, 0.3303]$ and $[0.2003, 0.3769]$.

Let the prior expectation of parameters equal to maximum likelihood estimators, thus $\mathbb{E}(\mu) = \nu = \hat{\mu} = 0.1178, \mathbb{E}(\tau) = \frac{\alpha}{\beta} = \frac{1}{\hat{\sigma}^2} = 11.99$ and the prior standard deviation of τ equals to a half of the confidence interval calculated in Section 2.3.3.1 with $std(\tau) = \frac{\sqrt{\alpha}}{\beta} = 0.0883$. Then the following values are considered, we have $\alpha = 18225, \beta = 1520, \nu = 0.1178$ and $k = 1$.

2.5.3.3 Calculation of the posterior estimator by MCMC

The two-dimension random walk Metropolis-Hastings algorithm is here used to generate the sample of 2 parameters with a two-dimensional Gaussian proposal distribution. Suppose we want to generate the samples of two parameters Y and Z , note the i^{th} generated parameters are y_i and z_i , the acceptance rate ζ is the same as in Equation (2.57), but replace the one-parameter distribution by a joint distribution of two parameters.

Generation algorithm

Here is the algorithm for using the MCMC :

1. Choose initial (y_0, z_0) for μ, σ from MLE estimators
2. Generate (y_{i+1}, z_{i+1}) from (y_i, z_i) with the proposed distribution corresponding to a two-dimension Gaussian distribution

3. Generate *r.v.* $U \sim Unif[0, 1]$,
 if $u > \zeta$
 accept (y_{i+1}, z_{i+1})
 else take $(y_{i+1}, z_{i+1}) = (y_i, z_i)$
4. Repeat Step 2~3 n times
5. Take the samples from $n/10$ to n times
5. end

Posterior distributions of μ and σ are sampled by Metropolis-Hastings of MCMC in Figure 2.16. The estimators are the mean values of sample, $\tilde{\mu} = 0.1146$ and $\tilde{\tau} = 11.4232$, thus $\tilde{\sigma} = 0.2959$.

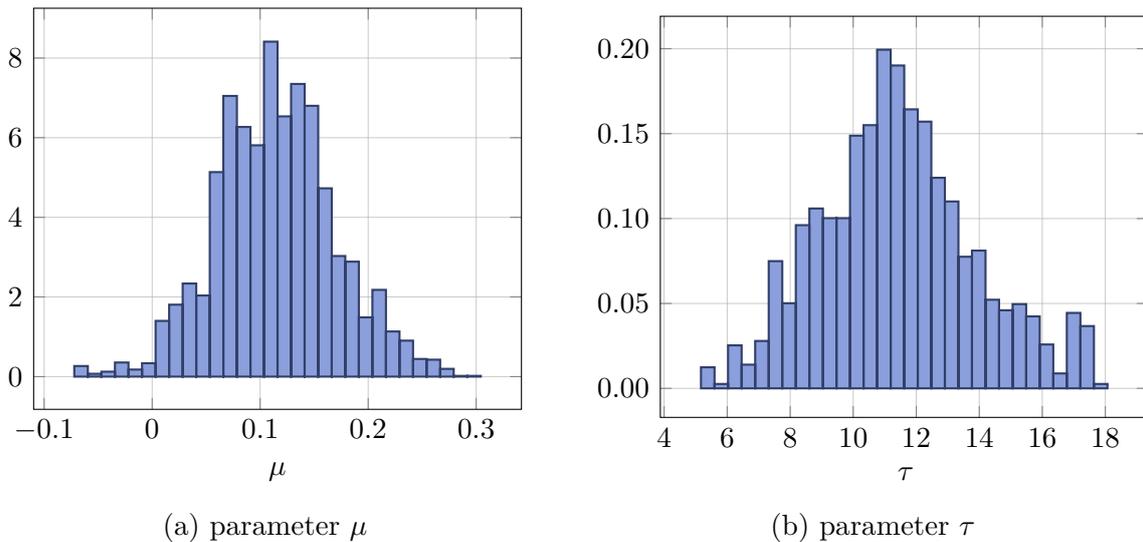


FIGURE 2.16 – Histogram of posterior distribution of parameters with normal-inverse gamma prior

2.5.3.3.1 Uniform prior and calculation of the posterior estimator by MCMC

According to the expert opinions in the company Renault, the prior information was taken as the uniform distribution on the confidence interval calculated with MLE. With the observations \underline{N}^f shown in Figure 2.3, the Bayesian estimation is done. The maximum likelihood estimators are $\hat{\mu} = 0.1178$ and $\hat{\sigma} = 0.2886$, their 95% confidence intervals are $[-0.0947, 0.3303]$ and $[0.2003, 0.3769]$ respectively. The prior distribution of μ and σ is taken as the uniform distribution in the 95% confidence interval.

The MH algorithm according to MCMC is used to sample the posterior distribution of parameters as in Figure 2.17. The estimates under square loss function are $\tilde{\mu} = 0.1180$ and $\tilde{\sigma} = 0.2833$.

2.5.4 Comparison among the different Bayesian estimation methods

In the Bayesian theorem, the choice of the prior distribution and its parameters play a vital role. In our study, the prior distribution parameter values are defined according

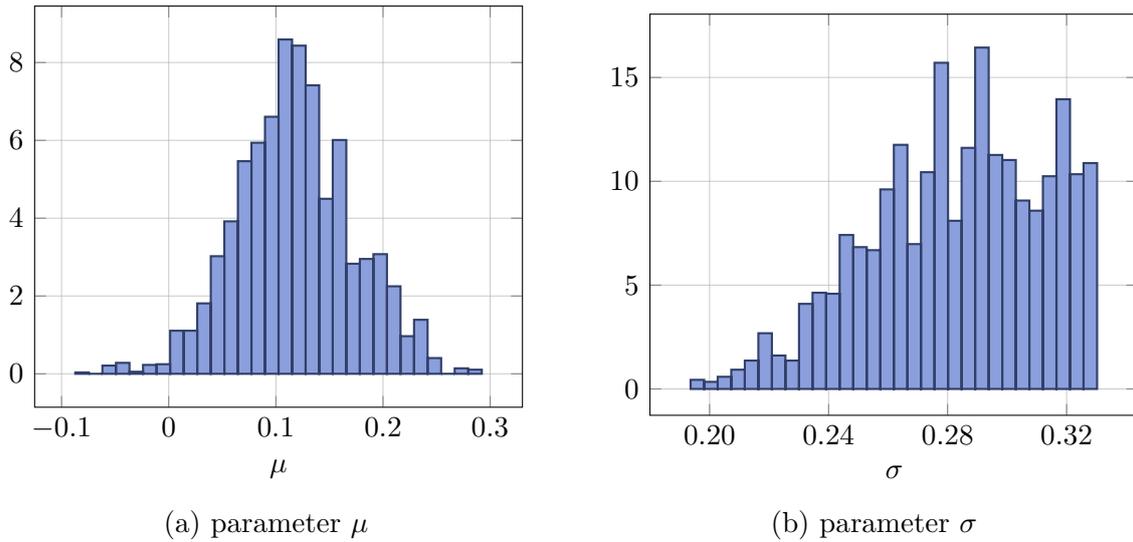


FIGURE 2.17 – Histogram of posterior distribution of parameters with uniform prior

to the results of MLE estimation. The PDF of experimental fatigue lifetime estimated by different Bayesian estimation methods are compared with that estimated by MLE estimation in Figure 2.18. The reliability functions and failure rate are compared in Figure 2.19. The Weibull and the Log-normal distributions are considered. The Bayesian estimation and MLE give similar results in terms of PDF, reliability and failure rate.

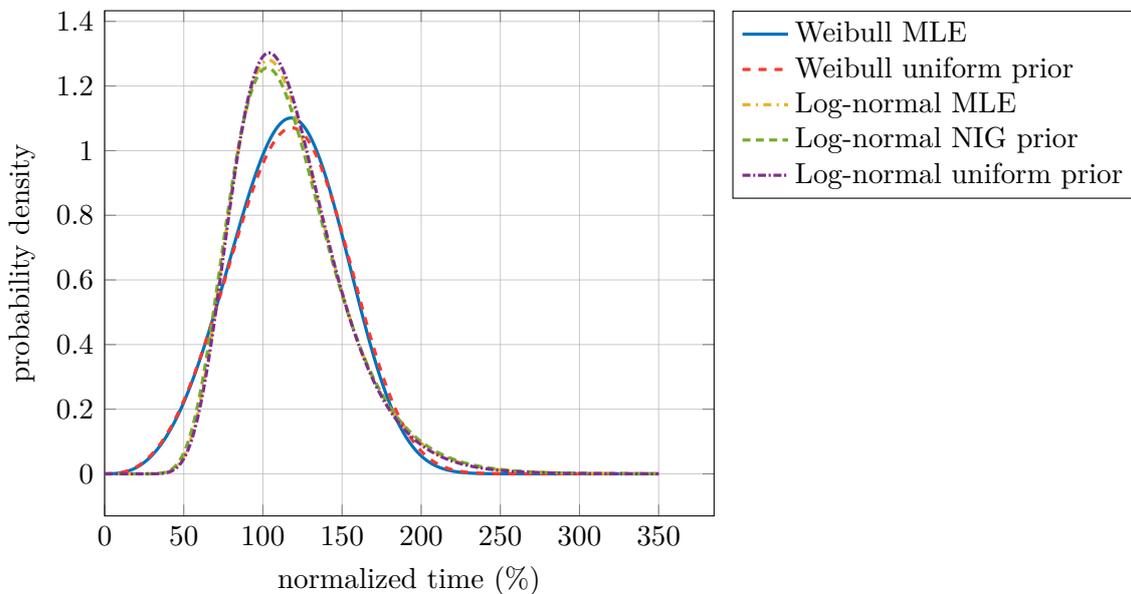


FIGURE 2.18 – Comparison of PDF estimated by MLE with different estimation methods for Weibull and Log-normal distributions

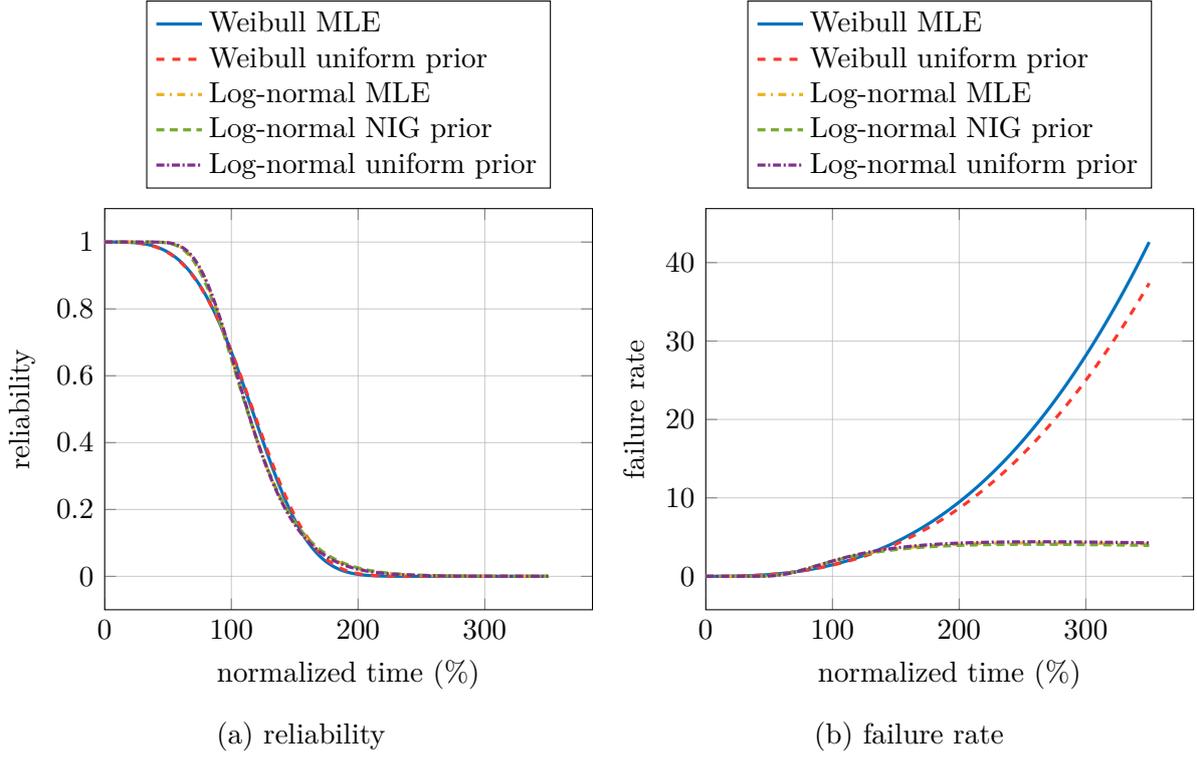


FIGURE 2.19 – Comparison of reliability function and failure rate obtained by MLE and different estimation methods for Weibull and Log-normal distributions

2.6 Estimation with multinomial distribution for grouped data

2.6.1 Definition

The multinomial distribution can be used to model the probability of counts for each event n_1, \dots, n_k with k the number of events. The total count is denoted as n , thus $n_1 + \dots + n_k = n$. As for grouped data, the fatigue lifetime for one interval can be considered as an event. The probability for a fatigue lifetime included in the i^{th} interval is p_i with $0 \leq i \leq k$, this can be modeled by a multinomial distribution.

The probability mass function (PMF) of a multinomial distribution is defined by

$$\begin{aligned}
 p(n_1, n_2, \dots, n_k | p_1, p_2, \dots, p_{m+1}, n) &= \mathbb{P}(N_1 = n_1, N_2 = n_2, \dots, N_{m+1} = n_{m+1}) \\
 &= \frac{\Gamma(n+1)}{\prod_{i=1}^{m+1} \Gamma(n_i+1)} \prod_{i=1}^{m+1} p_i^{n_i} \\
 &= \frac{n!}{n_1! \dots n_k!} \prod_{i=1}^{m+1} p_i^{n_i}
 \end{aligned} \tag{2.69}$$

where $p_i \in [0, 1], i = 1, \dots, k, P_1 + P_2 + \dots + P_k = 1$.

The expectation of \underline{N}^f is $\mathbb{E}(N_i) = np_i$, the variance is $\text{Var}(N_i) = np_i(1 - p_i)$.

In this model, the hypotheses of the fatigue lifetime continuous distribution type is no longer existing, the data are directly described in the form of discrete distribution law.

2.6.2 Conjugate prior and posterior distribution

The Bayesian estimation is done for multinomial distribution, whose conjugate prior distribution is the Dirichlet distribution $D(\alpha)$ [77].

Let us consider a random vector $\underline{P} = (P_1, P_2, \dots, P_k) \sim D(\alpha)$, the PDF is expressed as

$$\begin{aligned} p(p_1, \dots, p_k | \alpha_1, \dots, \alpha_k) &= \frac{1}{B(\underline{\alpha})} \prod_{i=1}^k p_i^{\alpha_i-1} \\ &= \frac{\Gamma(\sum_{i=1}^k \alpha_i)}{\prod_{i=1}^k \Gamma(\alpha_i)} \prod_{i=1}^k p_i^{\alpha_i-1}, \end{aligned} \quad (2.70)$$

where $\underline{\alpha} = (\alpha_1, \dots, \alpha_k)$, $\alpha_i > 0$, $B(\underline{\alpha})$ is the beta function, for example $B(x, y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}$. The expectation is $\mathbb{E}(P_i) = \frac{\alpha_i}{\alpha_0}$, where $\alpha_0 = \sum_{i=1}^k \alpha_i$.

Considering the Bayesian formula in Equation (2.48), and suppose the parameters to be estimated are $\underline{p} = (p_1, p_2, \dots, p_k)$. The observations follow the multinomial distribution $p(\underline{N} | \underline{p})$, the prior law is the Dirichlet distribution $D(\underline{\alpha})$, with parameters $\underline{\alpha} = (\alpha_1, \dots, \alpha_k)$. The posterior distribution $p(\underline{p} | \underline{N})$ is defined by

$$p(\underline{p} | \underline{N}) = \frac{\mathcal{L}(\underline{N} | \underline{p}) \pi_{pr}(\underline{p})}{p(\underline{N})} \sim D(\underline{N} + \underline{\alpha}). \quad (2.71)$$

The proof is as follows

$$\begin{aligned} p(\underline{p} | \underline{N}) &= \frac{\frac{\Gamma(n+1)}{\prod_{i=1}^k \Gamma(N_i+1)} \prod_{i=1}^k p_i^{N_i} \frac{1}{B(\underline{\alpha})} \prod_{i=1}^k p_i^{\alpha_i-1}}{\int \frac{\Gamma(n+1)}{\prod_{i=1}^k \Gamma(N_i+1)} \prod_{i=1}^k p_i^{N_i} \frac{1}{B(\underline{\alpha})} \prod_{i=1}^k p_i^{\alpha_i-1} dp} \\ &= \frac{\prod_{i=1}^k p_i^{N_i+\alpha_i-1}}{\int \prod_{i=1}^k p_i^{N_i+\alpha_i-1} dp}. \end{aligned} \quad (2.72)$$

The integral of the PMF over the value space equals to 1. For Dirichlet distribution $D(\underline{N} + \underline{\alpha})$, this integral is expressed as

$$\int_0^1 \dots \int_0^1 \frac{1}{B(\underline{N} + \underline{\alpha})} \prod_{i=1}^k p_i^{N_i+\alpha_i-1} dp_1 \dots dp_k = 1. \quad (2.73)$$

Thus

$$\int_0^1 \dots \int_0^1 \prod_{i=1}^k p_i^{N_i+\alpha_i-1} dp_1 \dots dp_k = B(\underline{N} + \underline{\alpha}). \quad (2.74)$$

Then the Equation (2.72) becomes

$$p(\underline{p} | \underline{N}) = \frac{1}{B(\underline{N} + \underline{\alpha})} \prod_{i=1}^k p_i^{N_i+\alpha_i-1}. \quad (2.75)$$

According to the PDF form of Dirichlet distribution in Equation (2.70), the posterior distribution is also a Dirichlet distribution $D(\underline{N}^f + \underline{\alpha})$. The expectation of posterior distribution in Equation (2.76) can be considered as the Bayesian posterior estimators of multinomial distribution under squared error loss function.

$$\hat{p}_i = \frac{N_i + \alpha_i}{\sum_{j=1}^k N_j + \alpha_j}. \quad (2.76)$$

2.6.2.1 Determination of prior distribution parameters value

Case I Firstly, the prior distribution is defined based on the hypotheses that the fatigue lifetime follow the Log-normal distribution or Weibull distribution, the results of MLE in Section 2.3.4 are used. The continuous PDF are divided by the intervals of fatigue lifetime observations. The integral of the PDF in each interval is considered to be the prior information of the p_i in multinomial distribution.

More precisely, suppose T denotes the fatigue lifetime, the parameter $\hat{\theta}$ is estimated by MLE, where $\hat{\theta}$ are the parameters for Log-normal or Weibull distribution or some other continuous distribution. Thus the PDF of fatigue lifetime is $f(t; \hat{\theta})$, the parameters of prior Dirichlet distribution are $\alpha_i = n \int_{t_{j-1}}^{t_j} f(t; \hat{\theta}) dt$. It corresponds to the probability for having a fatigue lifetime in the time interval $[t_{j-1}, t_j]$, with $j \in \llbracket 1, m+1 \rrbracket$ (see Figure 2.1), $t_0 = 0, t_{m+1} = +\infty$ and n is the total number of observations.

Case II Jeffrey prior is a non-informative prior [78]. For a distribution whose PDF is $f(x|\underline{\theta})$ and the parameter is $\underline{\theta} = (\theta_1, \theta_2, \dots, \theta_m)$, the definition of Jeffrey prior is expressed as

$$\pi_{pr}(\underline{\theta}) = |I(\underline{\theta})|^{1/2}, \quad (2.77)$$

where $I(\underline{\theta}) = -\mathbb{E}_{\underline{\theta}} \left[\frac{\partial^2 \log p}{\partial \theta^2} \right]$ is the Fisher information as in Equation (2.8).

The PMF of multinomial distribution in Equation (2.69) under logarithm is calculated

$$\log p(\underline{n}; \underline{p}, n) = \log \left(\frac{\Gamma(n+1)}{\prod_{i=1}^k \Gamma(n_i+1)} \right) + \sum_{i=1}^k n_i \log(p_i). \quad (2.78)$$

The Fisher information matrix of multinomial distribution is

$$I(\underline{\theta}) = \begin{cases} \mathbb{E}\left(\frac{N_i}{p_i^2}\right), & i = j \\ 0, & i \neq j \end{cases} \quad (2.79)$$

According to the property of multinomial distribution $\mathbb{E}(N_i) = np_i$, thus $\mathbb{E}\left(\frac{N_i}{p_i^2}\right) = \frac{n}{p_i}$

$$p(\underline{p}) \propto \prod_i^k p_i^{-1/2}. \quad (2.80)$$

Compared with the form of Equation (2.70), the conjugated prior of multinomial distribution with Jeffrey prior is defined by $\underline{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_k) = \left(\frac{1}{2}, \frac{1}{2}, \dots, \frac{1}{2}\right)$.

Case III Another natural non-informative prior distribution chooses the values $\underline{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_k) = (1, 1, \dots, 1)$.

Case IV With k observation intervals, a prior distribution is tested : $\underline{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_k) = (\frac{1}{k}, \frac{1}{k}, \dots, \frac{1}{k})$.

2.6.2.2 Results

Case I For the prior distribution defined above with Log-normal or Weibull law, the Bayesian estimators \hat{p}_i are taken as the expectation of posterior distribution. Finally, the CDFs and reliability estimated by Bayesian method are calculated and presented in Figure 2.20. It is observed that the CDF and the reliability modeled with multinomial distribution has the similar tendency as that estimated from Log-normal and Weibull distributions by MLE.

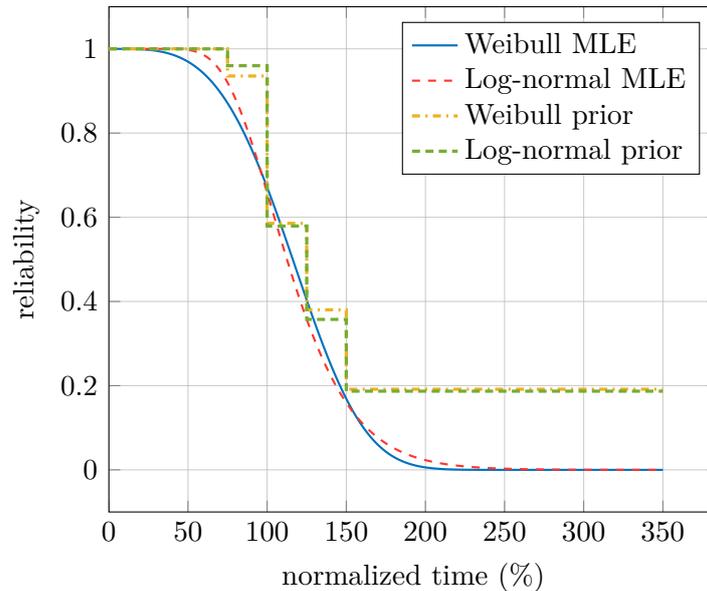


FIGURE 2.20 – Comparison of reliability obtained by MLE estimation of Weibull and Log-normal distributions with that obtained by Bayesian estimation of multinomial distribution

Case II, III, IV The parameters of prior distribution are $\underline{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_k) = (\frac{1}{2}, \frac{1}{2}, \dots, \frac{1}{2})$ in case II, $\underline{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_k) = (1, 1, \dots, 1)$ in case III, and $\underline{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_k) = (\frac{1}{k}, \frac{1}{k}, \dots, \frac{1}{k})$ in case IV. The CDFs and reliability function are calculated and presented in Figure 2.21 for case II, III and IV. This time with those non-informative prior distributions, the posterior multinomial distribution corresponds well that from MLE in the middle range of the CDF and reliability. In general, the reliability estimated by Bayesian method has the same tendency as that estimated by maximum likelihood method. Those 3 different non-informative prior distributions produce a reliability with a relatively large difference at the beginning, but becomes more and smaller after.

2.7 Simulated data estimation results

Here we used simulated data to verify the performance of different estimation methods and models, the MLE method and Bayes estimation of Weibull, Log-normal and multinomial distributions.

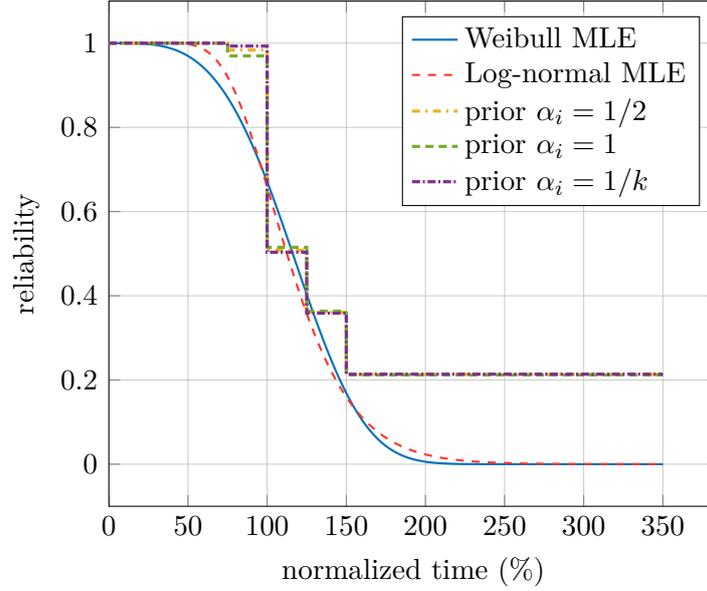


FIGURE 2.21 – Comparison of reliability obtained by MLE estimation of Weibull and Log-normal distributions with that obtained by Bayesian estimation of multinomial distribution with non-informative prior

2.7.1 Simulating method and simulated samples

The grouped data are generated from the discrete law multinomial distribution with the Log-normal distribution. According to the definition of multinomial distribution, it models the outcome of n experiments, such as rolling a k -sided die n times, the probability for having each side is p_i , with $i \in \llbracket 1, k \rrbracket$. So we generate the random numbers of multinomial distribution $f(n_1, n_2, \dots, n_k | p_1, p_2, \dots, p_k, n)$, where n is the total sample, and p_i is the probability for $r.v.$ N appearing in i^{th} interval with number n_i . The generating algorithm is as follows :

1. Define the value of p_i and $n_i = 0$
2. Generate a sample with size n of the $r.v.$ $X \sim Unif(0, 1)$
3. If $x_i \in [p_1 + \dots + p_i, p_1 + \dots + p_{i+1}]$, then $n_i = n_i + 1$
4. The sample \underline{n} is obtained.

The values of p_i are initialized with the help of a Log-normal distribution which is expected to be close to the real observed fatigue lifetime distribution. The parameters are chosen as $\mu = 0.1$ and $\sigma = 0.2$, and p_i equals to the integral of the Log-normal density function in the interval $[t_{i-1}, t_i]$, except the last one $p_m = 1 - \sum_{i=1}^{m-1} p_i$. The intervals are chosen as follows : $t = [0; 0.1; 0.2; 0.3; 0.4; 0.5; 0.6; 0.7; 0.8; 0.9; 1.0; 1.1; 1.2; 1.3; 1.4; 1.5; 1.6; 1.7; 1.8; 1.9; 2.0]$, 20 intervals in total.

The generation process is run for 100 times to evaluate the precision of the data. With those 100 generated samples, the estimation is done from simulated uncensored and grouped data, the estimated parameters are presented in Figure 2.22. The estimated $\hat{\mu}$ varies from 0.02 to 0.18 with the reference value 0.1, and the estimated $\hat{\sigma}$ varies from 0.15 to 0.25 with the reference value 0.2.

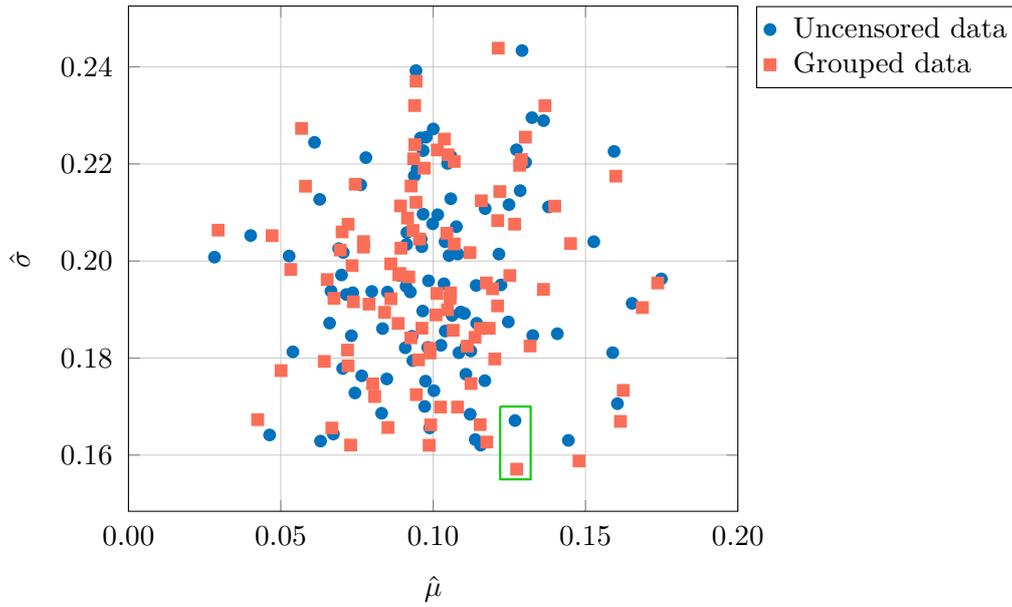


FIGURE 2.22 – Estimated Log-normal parameters with uncensored and grouped data

The difference between the simulated uncensored and grouped data are examined. With the pair of data marked by red rectangular in Figure 2.22 which shows relatively big difference between uncensored and grouped data, the estimated PDF by MLE method of exact and grouped data are presented in Figure 2.23. There is a difference observed on the peak of the probability density.

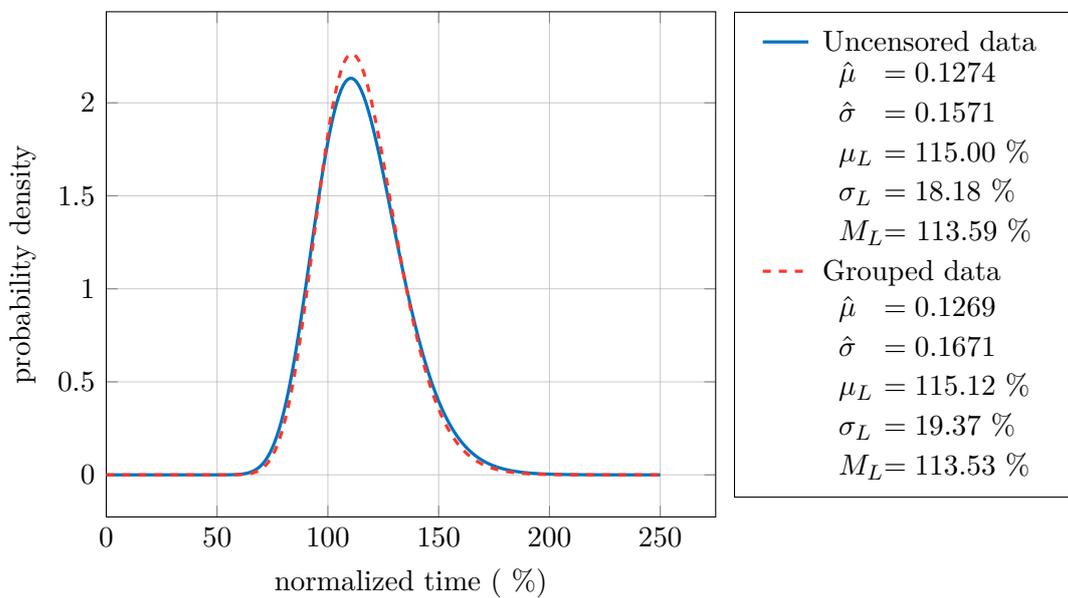


FIGURE 2.23 – Log-normal density distribution estimated by MLE from simulated uncensored and grouped data

For the study of simulated data, we want to choose a case where the grouped data do not lose much information compared to uncensored data. Thus the sample below has been chosen, the exact data sample is :

[0.6798; 0.7714; 0.7632; 0.7692; 0.8507; 0.8106; 0.8365; 0.8636; 0.8701; 0.8947; 0.9980; 0.9109; 0.9691; 0.9635; 0.9510; 1.0360; 1.0479; 1.0672; 1.0959; 1.0091; 1.0798; 1.0593; 1.0912; 1.0103; 1.0297; 1.1048; 1.1482; 1.1751; 1.1267; 1.1212; 1.1313; 1.1432; 1.1719; 1.1488; 1.2346; 1.2889; 1.2958; 1.2588; 1.2115; 1.2578; 1.3333; 1.3988; 1.3245; 1.3657; 1.4468; 1.4156; 1.4628; 1.5038; 1.6141; 1.7033]

then the grouped data sample is :

$t = [0; 0.1; 0.2; 0.3; 0.4; 0.5; 0.6; 0.7; 0.8; 0.9; 1.0; 1.1; 1.2; 1.3; 1.4; 1.5; 1.6; 1.7; 1.8; 1.9; 2.0];$

$n = [0; 0; 0; 0; 0; 0; 1; 3; 6; 5; 10; 9; 6; 4; 3; 1; 1; 1; 0; 0];$

The estimated PDF from uncensored and grouped data with MLE method are presented in Figure 2.24. It is calculated that the two PDF are very similar.

In addition, the empirical and theoretical Log-normal ($\mu = 0.1, \sigma = 0.2$) CDFs are compared in Figure 2.9. It is observed that the simulated sample submits generally to the theoretical distribution.

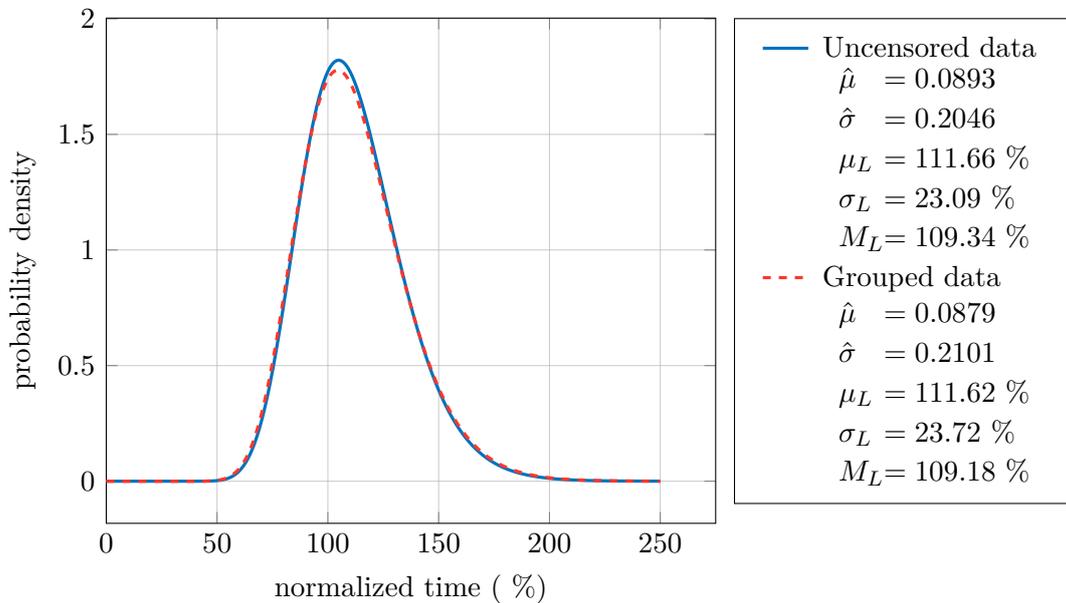


FIGURE 2.24 – Log-normal density distribution estimated by MLE from simulated uncensored and grouped data

2.7.2 Bayesian estimation for Log-normal distribution with grouped data

2.7.2.1 Normal-inverse-gamma prior

The prior distribution is supposed to be Normal-inverse-gamma distribution. The prior parameters are chosen according to the results of MLE. Then the Markov chain Monte-Carlo is used to sampling μ and $\tau = \frac{1}{\sigma}$ from the posterior distribution. The histogram of sample is presented in Figure 2.25. The mean value of the sample can be considered as the estimators under square error function $\tilde{\mu} = 0.0887$ and $\tilde{\tau} = 22.9725$, thus $\tilde{\sigma} = \sqrt{\frac{1}{\tilde{\tau}}} = 0.2086$.

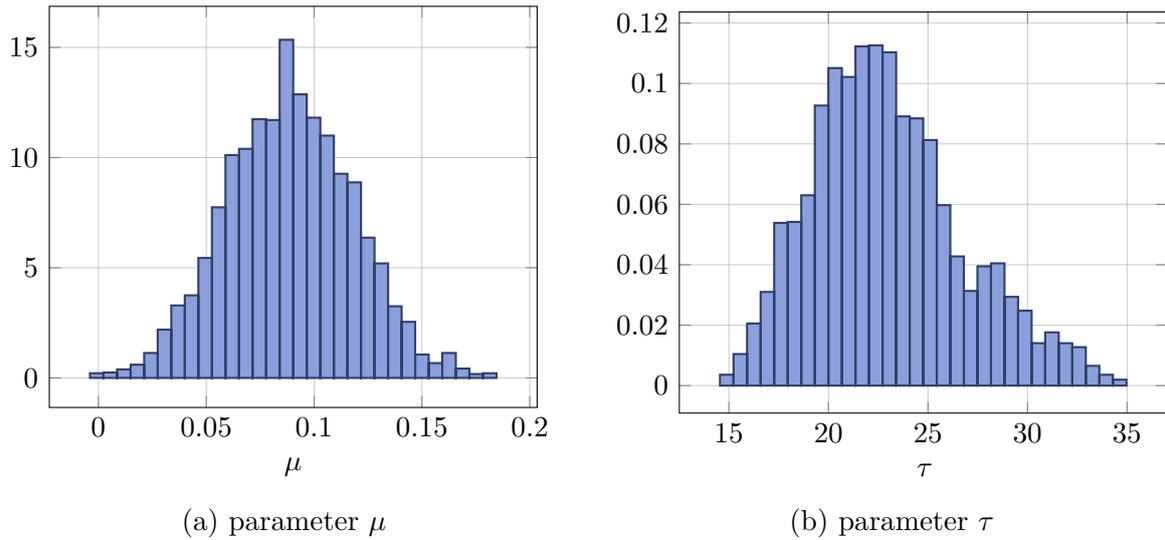


FIGURE 2.25 – Histogram of posterior Log-normal parameters with Normal-gamma prior for simulated data

2.7.2.2 Uniform prior

We can take the prior as the uniform distribution on the confidence interval calculated from MLE. Then the MCMC method is used to sample the posterior distribution, the histogram is presented in Figure 2.26. The estimators are the mean values of sample $\tilde{\mu} = 0.0876$, and $\tilde{\sigma} = 0.2262$.

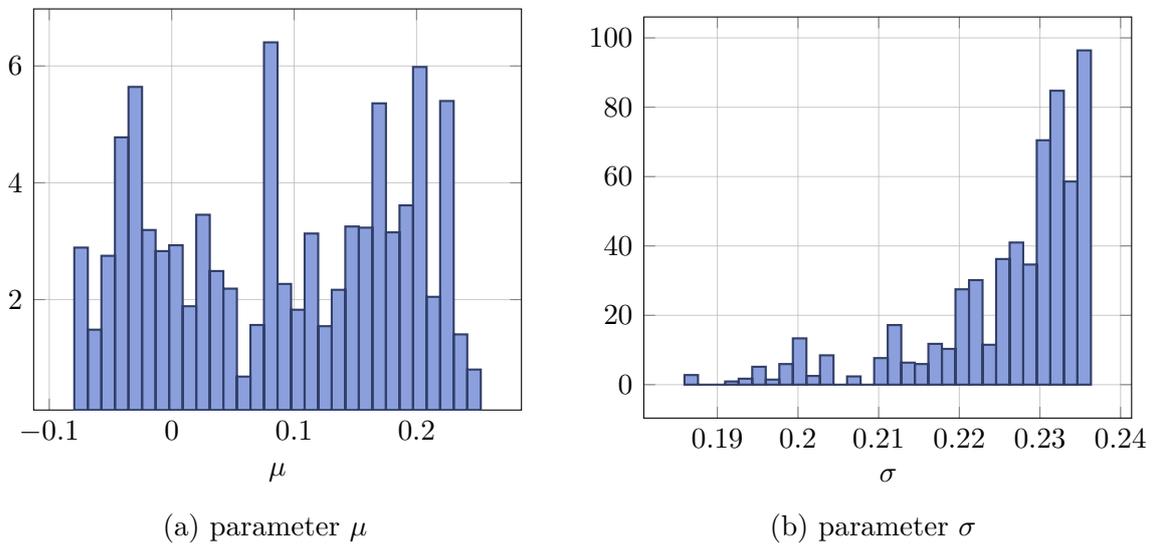


FIGURE 2.26 – Histogram of posterior Log-normal parameters with uniform prior for simulated data

2.7.2.3 Bayesian estimation for multinomial distribution

Three types of non-informative prior distributions are applied and their estimated cumulative distribution functions are compared with the theoretical distribution.

Comparison of estimation results The CDFs and the reliability estimated with different methods are compared in the Figure 2.27 to verify the performance of different methods. These methods are maximum likelihood estimation of Log-normal distribution, Bayesian estimation with Normal-inverse-gamma prior distribution and Bayesian estimation with uniform prior. Then for the estimation of multinomial distribution with three non-informative prior distribution, the parameters are $\underline{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_k) = (\frac{1}{2}, \frac{1}{2}, \dots, \frac{1}{2})$, $\underline{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_k) = (1, 1, \dots, 1)$, and $\underline{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_k) = (\frac{1}{k}, \frac{1}{k}, \dots, \frac{1}{k})$ respectively.

The Bayesian estimation for Log-normal distribution leads to a similar result as the MLE, because the prior parameters are defined with MLE results. For estimation of multinomial distribution, it is observed that the prior $\alpha_i = 1/k$ gives the closest estimation to the estimated continuous Log-normal distribution, which is better among the three non-informative prior distributions.

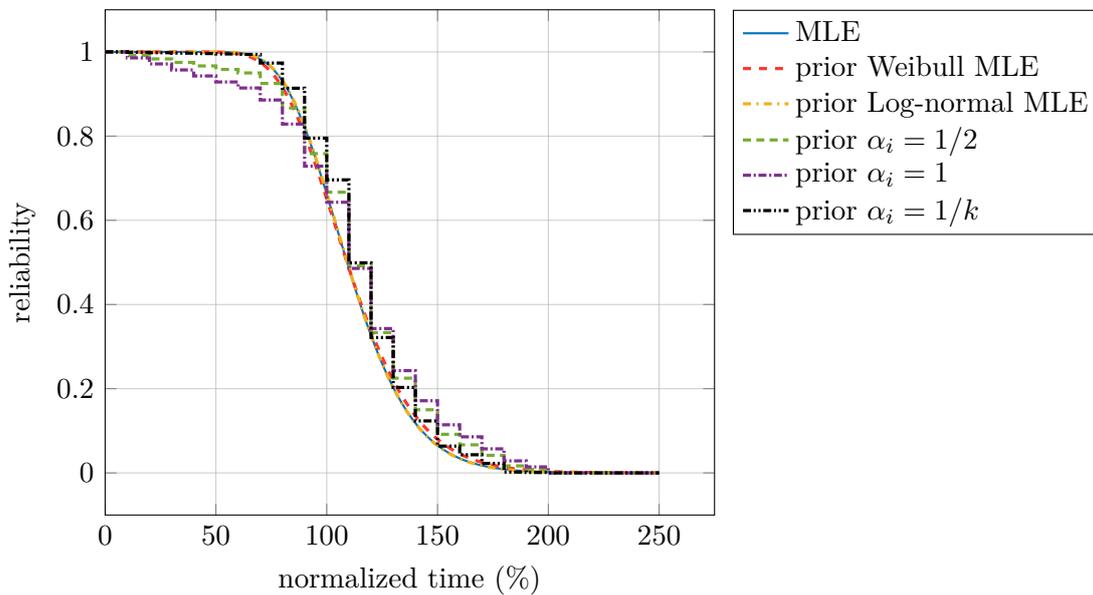


FIGURE 2.27 – Reliability estimated by different methods

2.8 Conclusion

In this chapter, the experimental fatigue lifetime grouped data observed in physical tests have been introduced and estimated by different numerical methods and models. Two types of methods have been proposed, the maximum likelihood estimation (MLE) and the Bayesian estimation method. Three distributions have been used to model the fatigue lifetime data : the continuous Log-normal and Weibull distributions, as well as the discrete multinomial distribution.

The MLE method estimates the parameters of a distribution from the experimental observations, the parameters to be estimated are determinist. The Bayesian estimation method estimates the parameters from the experimental observations and the prior distribution. The advantage of the Bayesian estimation is that it allows to take into consideration a prior information which can come from experts or previous studies. The Bayesian estimation reduces the 'bad' influence of inaccurately measured data, especially when

there are not many data. As for the prior distribution, the conjugated prior distributions are always considered as a good choice for the simplicity of computation, because the posterior distributions are the same type as the prior distribution. However, in the case of grouped data a tedious computation of posterior distribution is unavoidable. To do this computation, two approaches can be used : the approximation calculation and simulation method such as Markov Chain Monte Carlo. In our study, we use the method of MCMC, normally when the initial value and proposal distribution are well chosen, the MCMC method gives a good convergence performance. Besides, the non-informative prior distributions are also often adopted for Bayesian estimation when there is no information available.

As for different models of experimental fatigue lifetime, the continuous Log-normal and Weibull distributions are often used to model the lifetime of mechanical structure in the literature. In our study context, the Log-normal distribution is adopted according to the expert experiences in the Renault company. However, in the reliability analysis with those two distributions, the failure rate obtained with Log-normal distribution is not realistic because it gives a decreasing failure rate from the middle of lifetime. Normally the structure would have an increasing failure rate because of aging effect, but the Weibull distribution can describe this phenomenon. Thus if the estimated distribution will be used to analyse the failure rate, the Weibull distribution can be a better choice. The discrete multinomial distribution does not make hypothesis with the grouped data, and it describes the exact information contained in the data. But it is not yet mature to be integrated in the industrial application, further study is required.

With the results obtained in this chapter, we propose to choose the distribution model according to experts' opinions and to the objective of study. Bayesian estimation is strongly recommended when the number of data is small.

Chapitre 3

Stratégie d'identification des incertitudes liées à la courbe de Wöhler

Dans ce chapitre, une stratégie d'identification des incertitudes liées à la courbe de Wöhler est proposée, des nouvelles fonctions de coût sont également introduites pour mesurer la distance entre deux populations probabilistes. Une étude de convergence en fonction de la taille d'échantillon est menée sur les fonctions de coût, une étude de la précision d'identification du modèle est faite en s'appuyant sur un réseau de fonctions à la base radiale pour gagner en efficacité.

Contents

3.1 Introduction	69
3.2 Problématique	70
3.3 Approche analytique	70
3.3.1 Propagation directe/inverse avec un paramètre stochastique . .	71
3.3.2 Propagation directe avec deux paramètres stochastiques	71
3.4 Approche par simulation	72
3.4.1 Principe de l'approche par simulation	72
3.4.2 Construction des fonctions de coût	73
3.4.2.1 Comparaison paramétrique	73
3.4.2.2 Tests statistiques non paramétriques	74
3.4.2.3 Identification multi-sites	76
3.4.3 Méthode d'optimisation	77
3.5 Cas d'application - identification des deux paramètres de la courbe de Wöhler probabiliste	78
3.5.1 Données de référence	79
3.5.2 Identification des deux paramètres de la courbe de Wöhler P-S-N	79

3.5.3	Influence de la taille de l'échantillon sur la robustesse des fonctions de coût et de l'algorithme d'optimisation	81
3.5.3.1	Robustesse des fonctions coût	81
3.5.3.2	Comparaison des fonctions de coût	83
3.5.3.3	Identification par algorithme d'optimisation	86
3.5.4	Performance des fonctions de coût au voisinage du point de référence	86
3.6	Cas d'application - identification des trois paramètres de la courbe P-S-N	93
3.6.1	Identification à partir des données d'un seul cordon de soudure	93
3.6.2	Identification à partir des données de plusieurs cordons de soudure	95
3.6.3	Stratégie d'optimisation multi-niveaux adaptée	98
3.6.4	Identification des trois paramètres pour une autre référence . .	99
3.7	Précision des méthodologies d'identification	99
3.7.1	Génération des échantillons de μ_t et σ_t	102
3.7.1.1	Tirage des durées de vie groupées et estimation . . .	102
3.7.1.2	Tirage de μ_t et σ_t à partir de l'estimateur du maximum de vraisemblance	102
3.7.1.3	Comparaison des échantillons obtenus à partir des deux types de variabilité	104
3.7.2	Réseau de fonctions à base radiale	105
3.7.2.1	Réseau RBF exact	106
3.7.2.2	Réseau RBF approximant	107
3.7.3	Résultat de l'étude sur la précision des méthodologies d'identification	109
3.8	Conclusion	113

3.1 Introduction

L'objectif de ce chapitre est d'introduire la variabilité au niveau de la durée de vie calculée par la simulation numérique, afin de reproduire celle observée expérimentalement. L'hypothèse a été faite que cette variabilité est portée principalement par la courbe de Wöhler, autrement dit de multiples sources de variabilité sont intégrées dans la courbe de Wöhler. Ainsi, l'objectif devient d'identifier les incertitudes liées à la courbe de Wöhler probabiliste à partir des durées de vie expérimentales et celles calculées par la simulation numérique. En terme de dialogue, ce chapitre correspond à la partie B de la dialogue Figure 1.17 du chapitre 1. Le modèle probabiliste de la courbe de Wöhler a été introduit dans ce précédent chapitre. Dans le Chapitre 2, plusieurs méthodes d'estimation ont été proposées pour estimer la distribution des durées de vie expérimentales, en se basant sur le variabilité des données expérimentales. Les résultats d'estimation seront utilisés dans les méthodologies d'identification de la variabilité liée à la courbe de Wöhler dans ce chapitre. Deux approches d'identification ont été proposées : une approche analytique et une approche par simulation.

La méthodologie d'identification de la variabilité liée à la courbe Wöhler, qu'elle soit analytique, ou par simulation, est basée sur deux hypothèses. Premièrement, le modèle industriel déterministe actuel par éléments finis peut bien représenter les essais expérimentaux dans le cas nominal. Deuxièmement, la source principale d'incertitude est liée à la courbe de Wöhler.

L'approche analytique permet de construire une relation explicite entre la distribution des durées de vie et les paramètres stochastiques de la courbe de Wöhler pour certains choix de loi (voir [15]). Dans certains des cas traités ici, l'équation analytique existe. Pour l'approche analytique, la relation directe entre les paramètres estimés de la distribution de durée de vie et les paramètres probabilistes de la courbe de Wöhler est construite via l'équation analytique.

Pour l'approche par simulation, un problème inverse stochastique est construit. Lorsque le problème d'identification des paramètres est formulé dans un cadre probabiliste, il existe deux types de méthodes [79]. La première famille d'approches consiste à mesurer la différence entre les sorties observées et prédites du système par une fonction de coût et à essayer de trouver les paramètres du modèle stochastique qui minimisent cette différence [80, 81, 82]. La seconde famille d'approches considère que les données disponibles ne représentent que partiellement le modèle probabiliste des paramètres, et fournit un moyen de le mettre à jour progressivement lorsque des données supplémentaires sont disponibles [83]. Ces deux approches peuvent être combinées [84, 85]. Concernant la fonction de coût, Zabarar et Ganapathysubramanian [80] proposent une mesure de la distance sur les moments probabilistes (par exemple [81]) ou la densité de probabilité. Dans ce dernier cas, la fonction de coût est définie en terme de distribution cumulative inverse. Dans notre étude, 6 fonctions coût sont proposées pour mesurer l'écart sur la variabilité entre les durées de vie expérimentale et numérique. Ces nouvelles fonctions coût sont basées sur la comparaison paramétrique ou des tests statistiques. La performance des fonctions coût est étudiée avec des données simulées. Le comportement de la convergence en fonction de la taille de l'échantillon est analysé.

Pour mieux maîtriser le temps de calcul, Tran et Wildey [86] ont réalisé un travail en combinant un problème inverse stochastique et une technique de machine learning. Pour étudier la précision des méthodologies d'identification des paramètres, une étude par simulation a ici été faite. Un réseau à fonctions de base radiale a été construit comme une surface de réponse en prenant en compte des fonctions coût.

3.2 Problématique

Nous introduisons la problématique des travaux de ce chapitre basée sur le contexte présenté dans le chapitre 1. Une variabilité importante de la durée de vie est observée dans les essais synthétiques expérimentaux, mais la simulation numérique pratiquée actuellement par l'industrie mène à une durée de vie déterministe difficile à interpréter par rapport à cette variabilité expérimentale. L'objectif est donc de viser à obtenir une variabilité comparable pour simulation et les essais expérimentaux.

Dans cette étude, on suppose que le modèle numérique de l'essai synthétique développé chez Renault est représentatif. On suppose aussi que la variabilité observée dans les essais est portée par la courbe de Wöhler probabiliste. L'objectif des travaux est donc d'identifier les paramètres probabilistes de la courbe de Wöhler avec les durées de vie expérimentales et numériques. La distribution des durées de vie expérimentales a pu être estimée avec les méthodes proposées dans le Chapitre 2. Dans la suite de ce chapitre, deux approches pour identifier les paramètres probabilistes de la courbe de Wöhler sont proposées : l'approche analytique et l'approche par simulation.

3.3 Approche analytique

Avec la méthode de calcul de la durée vie, l'expression analytique de la durée de vie est déduite en fonction des paramètres. Elle permet de construire la relation directe entre les paramètres estimés de la distribution de durée de vie et les paramètres probabilistes de la courbe de Wöhler.

D'après la section 1.2.2, la loi de Basquin s'écrit $\log(N) = -b \log(S) + \log(a)$. Avec cette loi et l'utilisation de l'approche Rainflow, le dommage d est calculé à l'aide de l'Équation (3.1) pour un cycle de cahier des charges

$$d = \sum_{i=1}^r \frac{n_i}{N_i} = \sum_{i=1}^r \frac{n_i}{aS_i^{-b}} = \frac{1}{a} \sum_{i=1}^r n_i S_i^b, \quad (3.1)$$

et la durée de vie T est déduite avec

$$T = \frac{1}{n_c d} = \frac{a}{n_c} \left[\sum_{i=1}^r n_i S_i^b \right]^{-1}, \quad (3.2)$$

d est le dommage pour un cycle de chargement, n_c est le nombre de cycles du chargement dans un cahier des charges, la variable aléatoire T est la durée de vie rapportée à la durée du chargement du CDC.

3.3.1 Propagation directe/inverse avec un paramètre stochastique

Si le paramètre b est considéré déterministe et fixé, la partie $\sum_{i=1}^r [n_i S_i^b]^{-1}$ de l'équation 3.2 est une constante notée K . Ainsi, la durée de vie T suit la même type de loi que le paramètre a . Comme présenté au chapitre 2, la durée de vie est supposée suivre une loi Log-normale, notée par \mathcal{LN} . Nous notons que (μ_a, σ_a) les paramètres de la loi Log-normale pour le paramètre a , et (μ_t, σ_t) pour la loi Log-normale de la durée de vie. A partir de l'Équation (3.2), la relation analytique entre les hyperparamètres est

$$\begin{aligned}\mu_t &= \mu_a + \log(K) - \log(n_c), \\ \sigma_t &= \sigma_a.\end{aligned}\tag{3.3}$$

Ainsi la distribution de durée de vie T est déterminée à partir de la loi du paramètre a ,

$$T \sim \mathcal{LN}(\mu_a + \log(K) - \log(n_c), \sigma_a),$$

avec $K = \sum_{i=1}^r [n_i S_i^b]^{-1}$.

Dans le sens inverse, lorsque la distribution de la durée de vie expérimentale est obtenue par les méthodes présentées dans le Chapitre 2, les paramètres de la courbe de Wöhler peuvent être déduits analytiquement à partir de l'Équation (3.3) par

$$\begin{aligned}\mu_a &= \mu_t - \log(K) + \log(n_c), \\ \sigma_a &= \sigma_t.\end{aligned}\tag{3.4}$$

Ainsi, les paramètres stochastiques de la courbe de Wöhler sont obtenus.

3.3.2 Propagation directe avec deux paramètres stochastiques

Si les paramètres a et b sont tous deux considérés comme des variables aléatoires indépendantes, nous notons la densité de probabilité pour chacun des deux paramètres $f_A(a)$ et $f_B(b)$ respectivement. Selon la loi des fonctions des variables aléatoires, la densité de probabilité de la durée de vie s'écrit alors

$$f_T(t) = \int_0^\infty f_A(n_c \sum_{i=1}^r n_i S_i^v) f_B(v) (n_c \sum_{i=1}^r n_i S_i^v) dv.\tag{3.5}$$

Démonstration :

Supposons $k(b) = n_c \sum_{i=1}^r [n_i S_i^b]^{-1}$, les fonctions

$$g(a, b) = \begin{cases} t = g_1(a, b) = ak(b) \\ v = g_2(a, b) = a \end{cases}\tag{3.6}$$

$$g^{-1}(t, v) = \begin{cases} a = v \\ b = k^{-1}(t/v) \end{cases}\tag{3.7}$$

Selon la loi de fonction des variables aléatoires, on a

$$f_{T,V}(t, v) = f_{A,B}(g^{-1}(t, v)) |DJ_{g^{-1}}(t, v)| = f_{A,B}(v, k^{-1}(t/v)) \frac{v}{|k'(k^{-1}(t/v))|} \quad (3.8)$$

Le déterminant de la matrice Jacobienne de fonction $g(a, b)$ est

$$DJ_{g(a,b)} = \det \begin{pmatrix} k(b) & ak'(b) \\ 1 & 0 \end{pmatrix} = ak'(b).$$

Pour la fonction inverse $g^{-1}(a, b)$, le déterminant est

$$DJ_{g^{-1}(a,b)} = \frac{1}{DJ_{g}og^{-1}(t, v)} = -\frac{1}{vk'(k^{-1}(t/v))},$$

où $k'(b) = \frac{-n_c \sum n_i S_i^b \ln S_i}{(n_c \sum n_i S_i^b)^2}$. Donc l'Équation (3.5) est obtenue.

En général, quand on a une densité de probabilité $f_T(t)$, on vérifie si l'intégrale de la fonction est égale à 1 sur le support $[0, +\infty)$ (Annexe E).

La densité de probabilité de la durée de vie pour la courbe de Wöhler à deux paramètres stochastiques est obtenue dans une voie directe. Pour la voie inverse, la déduction du modèle des paramètres a et b de la loi de Basquin analytiquement, en sachant la loi de durée de vie expérimentale, est également un aspect intéressant à étudier.

3.4 Approche par simulation

3.4.1 Principe de l'approche par simulation

L'objectif de la méthodologie proposée est d'identifier le modèle de la courbe S-N probabiliste à partir de la durée de vie expérimentale T^{exp} et de la durée de vie issue du calcul par éléments finis T^{EF} pour les cordons de soudure du composant automobile. On s'attend à ce que la durée de vie T^{EF} puisse représenter la variabilité de la durée de vie T^{exp} .

L'idée principale de l'approche par simulation est de construire des fonctions coût qui mesurent l'écart sur la dispersion entre les durées de vie T^{EF} et T^{exp} . Ce sont des fonctions des paramètres probabilistes (μ_a, σ_a) de la courbe de Wöhler. Les fonctions coût sont minimisées par rapport aux paramètres (μ_a, σ_a) pour identifier le modèle de la courbe de Wöhler qui permet de réduire l'écart sur la dispersion entre les deux durées de vie. La fonction de coût est donc cruciale pour la performance de la méthode, car elle mesure la différence entre les dispersions des données expérimentales et numériques de durée de vie.

La dispersion de la durée de vie expérimentale a été estimée dans le Chapitre 2. Par ailleurs, la durée de vie dite éléments finis T^{EF} est obtenue à l'aide des contraintes éléments finis et le post-traitement utilisant le modèle de courbe S-N probabiliste présenté dans le Chapitre 2. Les hyperparamètres de ce modèle à identifier sont $\theta = (\mu_a, \sigma_a)$. Dans le processus d'identification, d'une part, la contrainte temporelle est comptée par la technique Rainflow pour obtenir l'amplitude de la contrainte S_i et le nombre de cycle correspondant n_i , d'autre part, l'échantillon aléatoire de a est généré par une simulation de Monte-Carlo. L'échantillon de la durée de vie dite éléments finis est ensuite calculé à

partir de S_i et n_i en utilisant l'échantillon de a . Ensuite, les fonctions de coût $CF(\theta)$ sont construites avec l'échantillon généré ou la distribution estimée d'échantillon. La valeur de θ qui minimise la fonction de coût est θ estimé. Selon la fonction de coût, les exigences relatives à la quantité des données expérimentales et à la taille de l'échantillon numérique pour l'identification peuvent être différentes. Nous proposerons plusieurs types de fonctions de coût statistiques et étudierons leurs performances par rapport aux données expérimentales groupées de durée de vie.

La stratégie proposée est suffisamment générale pour être appliquée à un cadre large de problèmes d'identification et de problèmes inverses stochastiques indépendamment des modèles probabilistes et du nombre d'hyperparamètres.

3.4.2 Construction des fonctions de coût

Une fois que les informations statistiques des données expérimentales sur la durée de vie sont extraites (chapitre 2), les hyperparamètres du modèle de la courbe S-N stochastique peuvent être identifiés par la minimisation d'une fonction de coût qui mesure la distance entre la dispersion des données expérimentales et celle de la simulation numérique (Équation (3.9)). Comme mentionné précédemment, le choix de la fonction de coût $CF(\theta)$ est cruciale. Comme la densité de probabilité des durées de vie expérimentales n'est pas disponible en raison des données limitées, les approches standard qui utilisent la distance de Kullback-Leibler ne sont pas adaptées à notre travail [87]. Par conséquent, deux types d'approches alternatives sont proposés pour construire une fonction de coût mesurant la distance entre essai et modèle. Une comparaison paramétrique et l'utilisation de tests statistiques non-paramétriques sont proposés.

$$\hat{\theta} = \arg \min_{\theta} CF(\theta) \quad (3.9)$$

Cette identification est tout d'abord présentée à partir des durées de vie expérimentales sur un cordon de soudure de la structure, puis généralisée à la prise en compte de plusieurs cordons de soudure situés dans différentes zones de la structure.

3.4.2.1 Comparaison paramétrique

La durée de vie est censée suivre une loi Log-normale (section 2.1). L'estimation de la loi de Log-normale est effectuée d'une part pour les échantillons de durée de vie en fatigue expérimentale et d'autre part pour leur équivalent éléments finis à l'aide des approches du cumul du dommage proposées au chapitre 1. Ainsi, les paramètres de la loi Log-normale pour la durée de vie expérimentale $(\mu_T^{exp}, \sigma_T^{exp})$ et les paramètres de la loi Log-normale pour la durée de vie dite éléments finis $(\mu_T^{EF}, \sigma_T^{EF})$ sont obtenus. Ensuite, la différence entre $(\mu_T^{exp}, \sigma_T^{exp})$ et $(\mu_T^{EF}, \sigma_T^{EF})$ peut être utilisée pour définir une fonction de coût. Étant donné que les amplitudes des deux paramètres μ et σ sont différentes, elles sont pondérées dans la formulation de la fonction de coût en divisant par μ_T^{exp2} et σ_T^{exp2}

$$CF1(\theta) = \frac{(\mu_T^{exp} - \mu_T^{EF})^2}{\mu_T^{exp2}} + \frac{(\sigma_T^{exp} - \sigma_T^{EF})^2}{\sigma_T^{exp2}}. \quad (3.10)$$

3.4.2.2 Tests statistiques non paramétriques

Le deuxième type de fonctions de coût étudié ici est basé sur des tests statistiques non paramétriques, qui s'appuient sur un test d'adéquation à un échantillon ou un test à deux échantillons. Nous considérons le test χ^2 , le test de Kolmogorov-Smirnov à un échantillon et le test de Cramér-von Mises à un échantillon. Les tests à deux échantillons tels que le test de Kolmogorob-Smirnov à deux échantillons et le test de Cramér-von Mises à deux échantillons sont également utilisés. Les détails de ces tests statistiques peuvent être trouvés dans [88] Chapitre 14 et sont brièvement rappelés ci-dessous.

Dans notre travail, nous considérons que la distribution Log-normale estimée à partir de l'échantillon de durée de vie expérimentale T^{exp} est une distribution connue pour la structure considérée, suite à une étape d'estimation telle que présentée dans le chapitre 2. L'échantillon de la durée de vie dite éléments finis T^{EF} est généré par une simulation de Monte-Carlo. Habituellement, les tests statistiques peuvent être utilisés pour décider si un échantillon suit une distribution donnée, ou si deux échantillons suivent la même distribution. Cette décision dépend de la distance entre la distribution et un échantillon ou entre deux échantillons. Elle s'appuie sur la comparaison entre la statistique de test et une valeur critique pour un niveau significatif, où la statistique de test est déterminée à partir d'observations. Ainsi, par exemple, pour un test d'adéquation à un échantillon, la statistique de test peut être considérée comme une mesure de la distance entre l'échantillon de durée de vie en fatigue dite éléments finis généré et la distribution de la durée de vie expérimentale. Nous proposons donc d'utiliser la statistique de test comme une fonction de coût.

3.4.2.2.1 Tests d'adéquation à une loi

Ces tests d'hypothèses statistiques visent à vérifier si un échantillon, correspondant ici à des données de durée de vie dite éléments finis, suit une loi de distribution connue $F_0(x)$, ici la distribution expérimentale estimée de la durée de vie. La fonction de distribution empirique de l'échantillon est notée $F_n^*(x)$, et est définie par

$$F_n^*(x) = k/n$$

où k est le nombre d'observations inférieures ou égales à x , et n est la taille de l'échantillon. L'hypothèse nulle des tests peut être $H_0 : F_n^*(x) = F_0(x), \forall x \in \mathbb{R}$, versus $H_1 : F_n^*(x) \neq F_0(x)$ pour $x \in \mathbb{R}$. La décision entre H_0 ou H_1 est faite par la comparaison entre la statistique calculée à partir des échantillons et une valeur critique. La statistique peut être considérée comme une mesure entre deux distributions. Trois tests d'adéquation sont maintenant présentés.

Test de χ^2 Les observations sont divisées en m classes selon la grandeur des données, avec des probabilités correspondantes p_1, p_2, \dots, p_m pour la distribution connue. Il y a n_1, n_2, \dots, n_m observations pour les m classes dans l'échantillon, $n = \sum_{i=1}^m n_i$ est le nombre total d'observations. L'objectif est de déterminer s'il existe une différence statistiquement significative entre les fréquences attendues et les fréquences observées dans les classes. La distribution de la statistique de test D^2 se rapproche de la distribution χ^2 de façon

asymptotique avec $m-1$ degrés de liberté lorsque l'hypothèse nulle est valide [88]. La statistique D^2 s'exprime comme suit,

$$D^2 = \sum_{i=1}^m \frac{(n_i - np_i)^2}{np_i}. \quad (3.11)$$

Pour décider si l'échantillon suit la loi visée, la statistique D^2 est comparée à une valeur critique sous un certain niveau de signification α pour décider d'accepter ou de rejeter l'hypothèse nulle H_0 . Comme D^2 mesure la distance entre la distribution connue F_0 et la distribution empirique de l'échantillon $F_n^*(x)$, cette statistique peut être considérée comme une fonction de coût : $CF2(\theta) = D^2$.

Test à un échantillon de Kolmogorov-Smirnov Supposons que F_n^* soit la fonction de distribution empirique des données expérimentales, la statistique de test D_n est construite sur la norme infinie de la différence entre la fonction de distribution empirique F_n^* et la fonction de la distribution F_0

$$D_n = \sup_{x \in \mathbb{R}} |F_n^*(x) - F_0(x)|. \quad (3.12)$$

Elle permet de constituer une fonction de coût : $CF3(\theta) = D_n$.

Test à un échantillon de Cramér-von Mises La statistique de test ω_n^2 présentée dans l'Équation (3.13) est construite sur la norme 2 de la distance entre la fonction de distribution théorique et la fonction de distribution empirique.

$$\omega_n^2 = \int_0^{+\infty} [F_n^*(x) - F_0(x)]^2 dF_0(x) \quad (3.13)$$

En supposant que $x_{(n)}$ est la statistique d'ordre de l'observation x (obtenue par tri croissant), et $u = F_0(x)$ de sorte que $u_{(n)} = F_0(x_{(n)})$, la forme de calcul pour l'intégrale dans l'Équation (3.13) est

$$\omega_n^2 = \frac{1}{12n} + \sum_{i=1}^n \left(u_{(i)} - \frac{2i-1}{2n} \right)^2. \quad (3.14)$$

Les détails du calcul sont présentés dans l'Annexe D. Il est alors possible d'utiliser cette statistique comme fonction coût : $CF4(\theta) = \omega_n^2$.

3.4.2.2.2 Test de comparaison de deux échantillons

Un autre type de test statistique non paramétrique peut être utilisé pour construire les fonctions de coût, le test à deux échantillons. Il est habituellement utilisé pour tester si deux échantillons suivent la même distribution. Dans le problème d'identification, il existe deux types de données : la durée de vie expérimentale T^{exp} et la durée de vie numérique T^{EF} générée par une courbe S-N probabiliste. Respectivement, ces deux échantillons ont des tailles n_1 et n_2 , leurs fonctions de distribution empirique correspondantes sont $F_{n_1}^*(x)$ et $F_{n_2}^*(x)$. Le test statistique à deux échantillons est alors réglé avec l'hypothèse nulle

$H_0 : F_{n_1}^*(x) = F_{n_2}^*(x)$. Il s'agit, ici encore, d'utiliser les statistiques associées aux tests pour construire la fonction coût.

Deux types de tests à deux échantillons sont utilisés :

Test à deux échantillons de Kolmogorov-Smirnov La différence entre les deux échantillons peut être évaluée par la statistique de Kolmogorov-Smirnov, la statistique de test D_{n_1, n_2} est considérée comme la fonction de coût

$$D_{n_1, n_2} = \sup_{x \in \mathbb{R}} |F_{n_1}^*(x) - F_{n_2}^*(x)|. \quad (3.15)$$

On peut alors définir la fonction coût par : $CF5(\theta) = D_{n_1, n_2}$.

Test à deux échantillons de Cramér-von Mises La comparaison des deux échantillons peut également être faite avec la statistique du test de Cramér-von Mises qui s'écrit comme suit

$$\omega_{n_1, n_2}^2 = \frac{n_1 n_2}{n_1 + n_2} \int_0^{+\infty} [F_{n_1}^*(x) - F_{n_2}^*(x)]^2 dH_{n_1 + n_2}(x), \quad (3.16)$$

où $H_{n_1 + n_2}(x) = \frac{n_1}{n_1 + n_2} F_{n_1}^*(x) + \frac{n_2}{n_1 + n_2} F_{n_2}^*(x)$. Le calcul numérique de l'intégrale de l'Équation (3.16) peut être trouvé dans [89].

Une autre fonction de coût peut être considérée : $CF6(\theta) = \omega_{n_1, n_2}^2$.

Dans la section suivante, les fonctions de coût sont appelées CF avec CF1 la comparaison paramétrique, CF2 la fonction de coût construite par le test de χ^2 , CF3 la statistique du test de Kolmogorov-Smirnov à un échantillon, CF4 la statistique du test de Cramér-von Mises, CF5 la statistique du test de Kolmogorov-Smirnov à deux échantillons et CF6 la statistique du test de Cramér-von Mises.

TABLEAU 3.1 – Fonctions coût proposées pour les méthodologies d'identification

Fonctions coût	Méthodes
CF1	Comparaison paramétrique
CF2	Statistique du test de χ^2
CF3	Statistique du test à un échantillon de Kolmogorov-Smirnov
CF4	Statistique du test à un échantillon de Cramér-von Mises
CF5	Statistique du test à deux échantillons de Kolmogorov-Smirnov
CF6	Statistique du test à deux échantillons de Cramér-von Mises

3.4.2.3 Identification multi-sites

Les fonctions de coût ont été initialement proposées pour traiter un seul joint de soudure. Lorsque les fonctions de coût sont optimisées pour déterminer les hyperparamètres de la courbe S-N probabiliste, les résultats sont donc obtenus pour chaque joint de soudure. Cependant, il peut être pertinent de considérer un unique modèle de courbe S-N pour tous les joints de soudure, car ils sont composés du même matériau. Il est possible d'adapter les fonctions coût précédentes pour tenir compte des données de plusieurs sites.

Pour cela, une fonction coût composite est proposée comme la somme des fonctions coût de chaque site.

3.4.3 Méthode d'optimisation

Une optimisation de la fonction de coût est nécessaire pour trouver son minimum. Les fonctions de coût peuvent avoir des minimums locaux. Par conséquent, les méthodes d'optimisation globale sont prises en considération. Plusieurs méthodes méta-heuristiques d'optimisation globale ont été évaluées : la méthode du recuit simulé, la méthode génétique et la méthode de l'essaim de particules. Finalement, la méthode de l'essaim de particules est choisie pour son efficacité sur tous les tests effectués. Une fois que l'optimisation globale a atteint son minimum, une optimisation locale basée sur une méthode de gradient est appliquée pour améliorer la précision de la solution.

Du fait du manque de régularité locale des fonctions coût, les méthodes d'optimisation exploratoires sont proposées pour les optimiser. Trois méthodes sont testées, l'algorithme de recuit simulé (RS), l'algorithme génétique (AG) et l'algorithme d'optimisation par essais particuliers (EP). Trois fonctions de coût CF1, CF3 et CF4 ont été minimisées par ces trois méthodes, dans le cas où les fonctions de coût sont construites avec des données simulées de taille 1000, les résultats d'optimisation et le temps du calcul sont présentés dans le Tableau 3.2. Avec les valeurs de référence $\mu_a = 20$ et $\sigma_a = 6$, les trois méthodes fournissent toujours des valeurs proches de μ_a et σ_a de référence.

Pour les méthodes AG et EP, les valeurs initiales sont définies aléatoirement dans l'intervalle $[0, 50]$ pour μ_a et $[0, 20]$ pour σ_a , et pour la méthode RS la valeur initiale est $[1, 1]$. En termes de temps d'optimisation et minimum trouvé des fonctions, la méthode EP est toujours plus rapide et arrive à déterminer la valeur la plus petite. La performance de l'algorithme d'optimisation peut dépendre des différents paramètres choisis dans l'algorithme et de la valeur initiale. Ici nous avons choisi les paramètres proposés par défaut. Cette comparaison simple (Tableau 3.2) montre que la méthode EP a une meilleure performance. Elle est donc retenue pour sa précision d'optimisation et son temps de calcul, ses grands principes sont donnés ci-dessous.

TABLEAU 3.2 – Résultats d'optimisation avec les méthodes RS, AG et EP

Fonction	Méthode	μ_a	σ_a	t(s)	Valeur de la fonction
CF1	RS	19.6673	5.9577	91.5588	0.0017
	AG	20.0633	6.0436	89.6785	$4.2867e^{-5}$
	EP	20.2672	5.8864	51.3006	$3.2246e^{-6}$
CF3	RS	19.9921	5.9705	83.1511	0.0206
	AG	20.0046	6.0597	415.0753	0.0206
	EP	19.7903	5.8627	74.4890	0.0110
CF4	RS	19.9733	6.2918	129.2343	0.0186
	AG	20.0576	5.8860	177.5414	0.0372
	EP	20.1283	5.8430	68.6807	0.0174

L'algorithme d'optimisation par essais particuliers a été proposée par Russel Ebe-

3.5. Cas d'application - identification des deux paramètres de la courbe de Wöhler probabiliste

rhart et James Kennedy [90]. La méthode était inspirée par le biomimétisme du déplacement d'un groupe d'oiseaux. Une introduction des paramètres dans l'algorithme est faite dans [90]. La position de la i^{ime} particule pour la d^{ime} composante de x au temps $t + 1$ est calculée à partir de la position actuelle $x_{id}(t)$, la vitesse actuelle $v_{id}(t)$, sa meilleure solution historique p_{best} et la meilleure solution historique de ses voisins g_{best} (voir Figure 3.1). Il y a différentes versions de l'algorithme, la plus utilisée actuellement est présentée par l'Équation (3.17)

$$\begin{aligned} v_{id}(t+1) &= wv_{id}(t) + y_1u_1(p_{best} - x_{id}(t)) + y_2u_2(g_{best} - x_{id}(t)) \\ x_{id}(t+1) &= x_{id}(t) + v_{id}(t+1) \end{aligned} \quad (3.17)$$

Le paramètre w est l'inertie sur la vitesse actuelle, les deux termes suivants modifient la vitesse pour tenir compte de la meilleure position de la particule et de ses voisins. u_1 et u_2 sont des variables aléatoires uniformes dans $[0, 1]$ rendant l'algorithme stochastique.

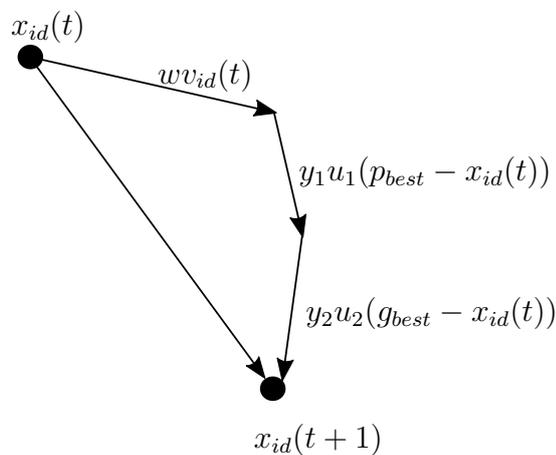


FIGURE 3.1 – Schéma de l'algorithme par essais individuelles

Plusieurs études ont été menées sur l'influence des paramètres dans l'algorithme [91]. Une proposition de [91] a été adoptée dans notre étude : $w = 0.7298, y_1 = 2.0434, y_2 = 0.9487$. Une fois un minimum global trouvé par l'algorithme exploratoire, une méthode de descente locale (points intérieurs) est appliquée pour faire une optimisation locale afin d'améliorer la précision de l'optimisation.

3.5 Cas d'application - identification des deux paramètres de la courbe de Wöhler probabiliste

La méthode d'identification de la courbe S-N probabiliste est ici validée pour des données de référence fabriquées par simulation. L'identification de la distribution Log-normale du paramètre a de la courbe S-N, $a \sim \mathcal{LN}(\mu_a, \sigma_a)$, est présentée. Ensuite, l'influence de la taille de l'échantillon sur les performances des fonctions de coût est étudiée. L'identification sera étendue à la section 3.6 au cas de trois paramètres considérés pour identifier (μ_a, σ_a) et le paramètre déterministe b correspondant à la pente de la loi de Basquin. Il

faut noter ici que les données de référence sont des données complètes et non des données groupées. Pour les fonctions de coût CF5 et CF6 construites avec deux échantillons, les études sont menées avec les données de référence. Pour l'application sur les données expérimentales groupées des études supplémentaires sont envisagées.

Dans un premier temps, le paramètre b est pris égale à la valeur de référence utilisée chez Renault. Les méthodologies sont utilisées pour identifier les deux paramètres (μ_a, σ_a) du modèle de la courbe de Wöhler P-S-N.

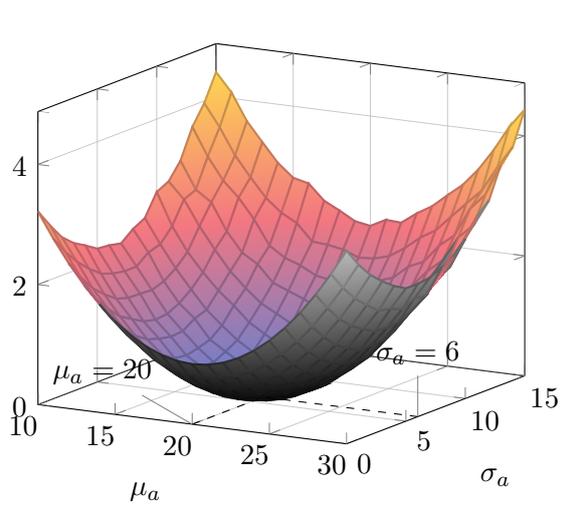
3.5.1 Données de référence

Comme mentionné précédemment, le paramètre a de la loi de Basquin est modélisé par une variable aléatoire suivant une distribution Log-normale avec les paramètres $\theta = (\mu_a, \sigma_a)$. La contrainte temporelle d'un joint de soudure obtenue par l'analyse par éléments finis est comptée par la méthode Rainflow pour obtenir la séquence d'amplitudes S et le nombre de cycles correspondant N . Un paramètre de référence θ^{ref} pour la courbe P-S-N est utilisé pour générer un échantillon de référence de durée de vie en fatigue T^{exp} par simulation de Monte Carlo représentant les données expérimentales de durée de vie en fatigue. Ensuite, le processus d'identification est utilisé pour vérifier si le paramètre identifié θ est proche de θ^{ref} . Dans l'étude suivante, nous prenons $\theta^{ref}=(20,6)$.

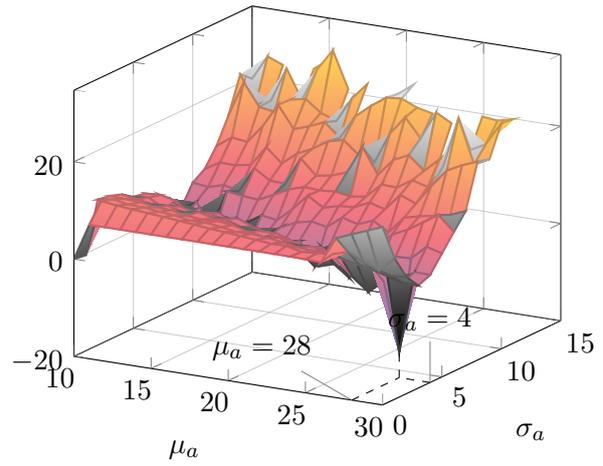
3.5.2 Identification des deux paramètres de la courbe de Wöhler P-S-N

Les différentes fonctions de coût mentionnées dans la section 3.4.2 sont appliquées avec des données de référence sur la durée de vie. La Figure 3.2 montre les valeurs des fonctions de coût dans l'espace des paramètres (μ_a, σ_a) . Les minimums des fonctions de coût sont mis en évidence. La Figure 3.2 montre que les minimums des fonctions de coût se situent autour de $(\mu_a, \sigma_a) = (20, 6)$ sauf pour la fonction de coût CF2. Cette fonction CF2 est représentée sur une échelle logarithmique en raison de sa grande plage de valeurs. En répétant sa minimisation plusieurs fois, les minimums de CF2 construite avec le test χ^2 ne sont pas stables et peuvent être éloignés de la valeur de référence. Ainsi, la fonction coût CF2 n'est pas considérée comme utilisable. Quant aux autres fonctions de coût, elles donnent une valeur minimale précise et semblent lisses avec une minimisation numérique. De plus, les fonctions de coût CF3 et CF5 ont des performances similaires puisqu'elles sont basées sur la même type de statistique, tout comme pour CF4 et CF6 pour la même raison. La méthode d'identification proposée pour la courbe S-N probabiliste est ici considérée comme validée car le paramètre de référence $\theta^{ref} = (20, 6)$ peut être trouvé avec une bonne précision (moins de 1% d'erreur).

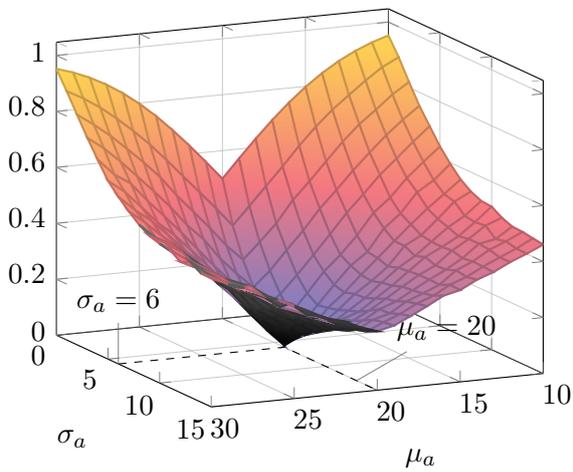
La forme des fonctions de coût observées dans la Figure 3.2 peut influencer la performance de l'algorithme d'optimisation. La taille de l'échantillon de durée de vie en fatigue dite éléments finis T^{EF} peut également influencer l'efficacité de l'identification.



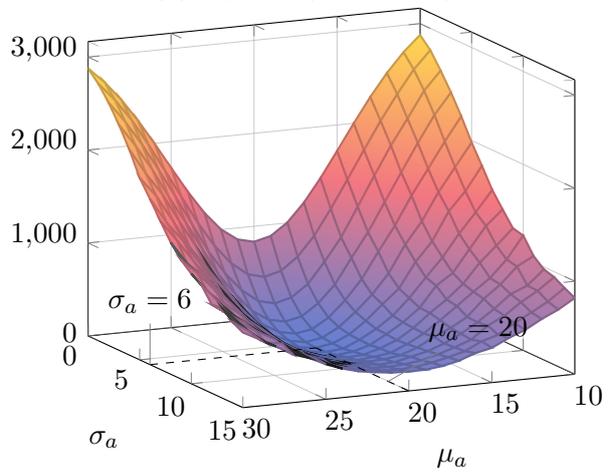
(a) CF1 (Comparaison paramétrique)



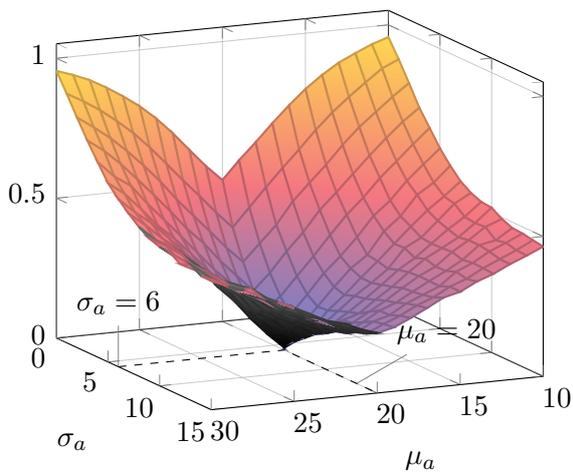
(b) logCF2 (test du χ^2)



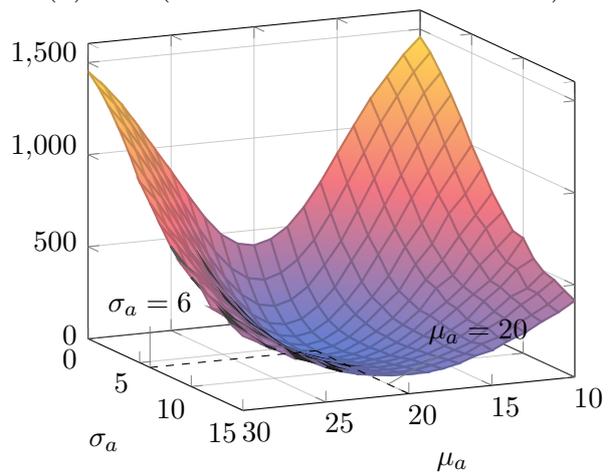
(c) CF3 (test à un échantillon de KS)



(d) CF4 (test à un échantillon de CVM)



(e) CF5 (test à deux échantillons de KS)



(f) CF6 (test à deux échantillons de CVM)

FIGURE 3.2 – Valeurs des six fonctions de coût dans l'espace des paramètres (μ_a, σ_a)

3.5.3 Influence de la taille de l'échantillon sur la robustesse des fonctions de coût et de l'algorithme d'optimisation

L'influence de la taille de l'échantillon est ici utilisée pour évaluer les fonctions coût et pour vérifier la convergence de la méthode d'identification. Deux parties de la robustesse peuvent être distinguées : l'influence sur la robustesse des fonctions de coût et l'influence sur l'algorithme d'optimisation.

3.5.3.1 Robustesse des fonctions coût

Tout d'abord, la convergence de la simulation de Monte Carlo est étudiée. Les six fonctions de coût sont calculées pour différentes tailles d'échantillons de durée de vie en fatigue dite éléments finis $n = \{10, 100, 1000, 10000\}$. Pour chaque cas, la simulation est effectuée pour plusieurs temps de répétition $n_r = \{1, 2, 3, \dots, 1000\}$. Ensuite, la valeur moyenne et l'écart-type des échantillons de la fonction de coût sont évalués en fonction de n_r . L'étude montre une convergence des fonctions de coût lorsque le temps de répétition atteint $n_r = 400$.

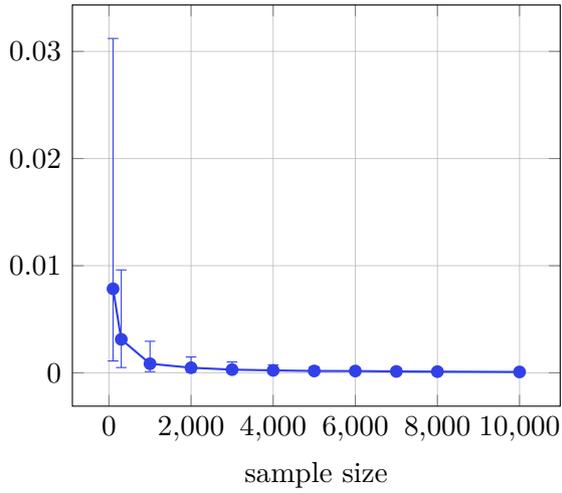
Lorsque la convergence est atteinte avec 400 simulations, les fonctions de coût sont calculées pour les différentes tailles d'échantillon afin d'évaluer la robustesse de ces fonctions. La taille de l'échantillon est fixée à $n = \{100, 300, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10000\}$. Ensuite, la valeur moyenne, les 10^{ème} et 90^{ème} percentiles des 400 simulations pour chaque taille d'échantillon sont calculés et présentés dans la Figure 3.3. Une bonne convergence est observée pour les fonctions de coût CF1, CF3 et CF5. Pour les fonctions de coût CF2, CF4 et CF6, une tendance de convergence n'est pas observée en termes de moyenne et de 10^{ème} et 90^{ème} percentiles, ce qui signifie que la taille de l'échantillon n'a pas d'effet sur la fonction coût et sa variabilité.

On observe également que pour les fonctions de coût CF1, CF2, CF4 et CF6, le 10^{ème} percentile est proche de la médiane mais le 90^{ème} percentile est loin de la médiane. cela signifie que la distribution est significativement asymétrique à gauche, CF3 est moins asymétrique.

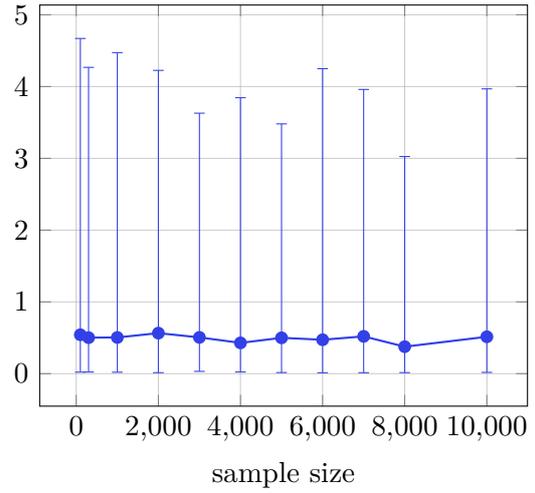
Pour une meilleure compréhension, la taille d'échantillon $n = 3000$ est choisie et l'histogramme des fonctions de coût pour 400 répétitions est présenté sur la Figure 3.4. Il est conforme que les quatre fonctions de coût CF1, CF2, CF4 et CF6 sont manifestement asymétriques à gauche. Les fonctions de coût CF3 et CF5 ont tendance à présenter une distribution davantage symétrique par rapport aux autres fonctions. Pour les fonctions de coût basées sur des statistiques, cela peut provenir des propriétés asymptotiques des statistiques de test.

En effet, les fonctions de coût CF2 sont basées sur la statistique du test χ^2 . Et les fonctions CF3 et CF5 sont basées sur les statistiques de test de Kolmogorov-Smirnov (KS).

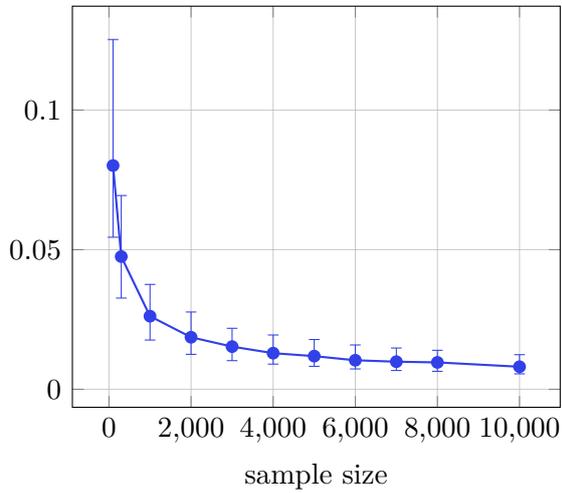
La statistique de test de Cramér-von Mises (CVM) est considérée pour CF4 et CF6. La distribution asymptotique de la statistique χ^2 est la distribution χ^2 , qui est asymétrique à gauche.



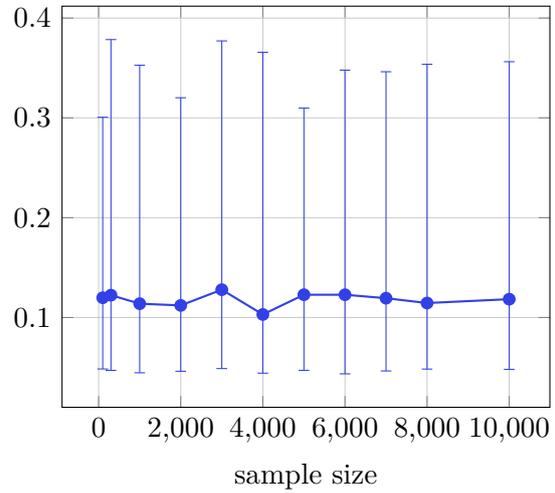
(a) CF1



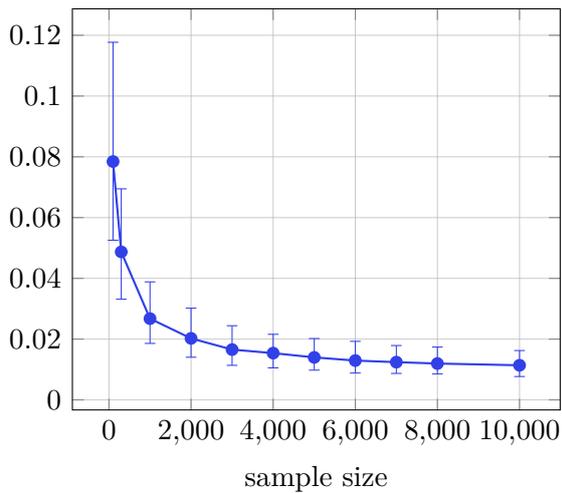
(b) CF2



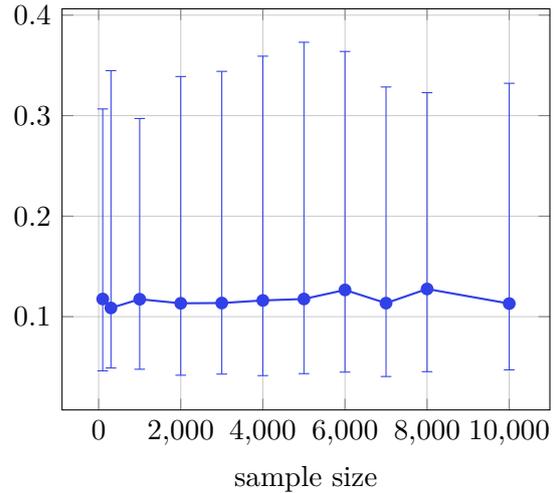
(c) CF3



(d) CF4



(e) CF5



(f) CF6

FIGURE 3.3 – Médiane, 10^{ème} et 90^{ème} percentiles des six fonctions de coût pour différentes tailles d'échantillon T^{EF}

La distribution asymptotique du test KS est le supremum du pont brownien, qui est une distribution similaire à une distribution normale. Pour le test CVM, il s'agit de la somme d'une distribution infinie de χ^2 , qui est donc asymétrique à gauche.

Comme il est mentionné ci-dessus à partir de la Figure 3.3, on observe que les valeurs des fonctions de coût CF4 et CF6 avec les paramètres de référence sont presque constantes pour différentes tailles d'échantillon. Les variabilités des fonctions de coût ne varient quasiment plus avec le changement des tailles d'échantillon. Un test de médiane est ici utilisé pour observer si les échantillons de fonction de coût correspondant à différentes tailles d'échantillon T^{EF} proviennent de la même population, et donc pour vérifier si les populations des fonctions coût CF4 et CF6 de petite taille d'échantillon sont représentatives. Plusieurs tests existent pour les tests de médiane, par exemple, le test des rangs signés de Wilcoxon pour un échantillon, le test du signe et le test du Wilcoxon-Mann-Whitney pour deux échantillons, et le test de Mood et le test de Kruskal-Wallis pour plusieurs échantillons [92]. Par rapport au test de Kruskal-Wallis, le test de Mood est moins précis mais moins sensible aux valeurs aberrantes. Le test de Kruskal-Wallis est donc choisi ici. De plus, les échantillons pour différentes tailles ont une distribution similaire, qui est une condition d'utilisation du test de Kruskal-Wallis. Le test de Kruskal-Wallis est effectué pour des échantillons de taille différente $n = \{100, 300, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10000\}$ de CF4 et CF6 et donne la valeur p de 0,5671 et 0,9708 respectivement.

La p-value est dans l'intervalle $[0, 1]$, elle représente, quand l'hypothèse nulle H_0 est vraie, la probabilité d'avoir un échantillon plus extrême que ce qu'on a observé. Si p-value est petite, la probabilité que H_0 a lieu est faible. Par conséquent, dans notre cas, lorsqu'elle est plus proche de 1, il est plus fiable de considérer que les échantillons proviennent de la même population. D'après les résultats, on peut considérer que les échantillons de la statistique CVM sont issus de la même population avec un niveau de signification de 50% pour CF4 et 90% pour CF6. On peut conclure que la taille de l'échantillon de T^{EF} n'influence pas significativement la performance des fonctions de coût CF4 et CF6.

Dans cette section, nous avons étudié l'influence de la taille de l'échantillon T^{FE} sur la robustesse des fonctions de coût, une convergence a été observée pour les fonctions de coût CF1, CF3, et CF5. En revanche, les fonctions de coût CF2, CF4 et CF6 ne semblent pas converger avec la taille de l'échantillon. Nous verrons ci-dessous que le comportement de CF2 est assez différent de celui de CF4 et CF6.

3.5.3.2 Comparaison des fonctions de coût

D'après les résultats de la section ci-dessus, les six fonctions de coût ont des performances différentes. Une comparaison est effectuée en combinant les valeurs des fonctions de coût de la Figure 3.2 et l'étude de la robustesse des six fonctions de coût de la Figure 3.3.

Pour les six fonctions de coût, ni les valeurs de la fonction de coût ni l'intervalle entre le 10^{ème} percentile et le 90^{ème} percentile ne peuvent être comparés directement. On souhaite également savoir si la dispersion observée sur la Figure 3.3 est importante par rapport à l'ordre de grandeur de la fonction de coût. Nous avons donc proposé de

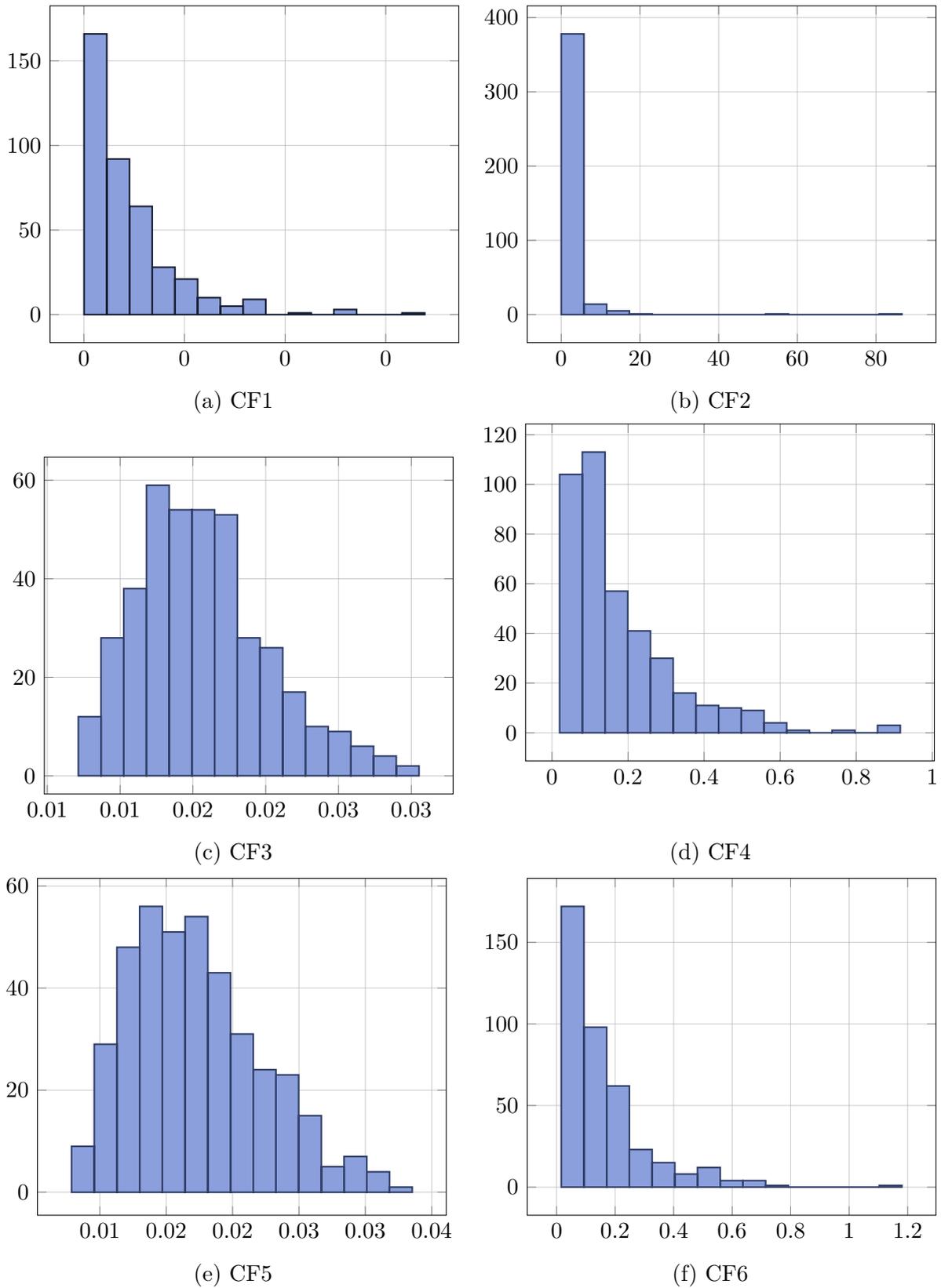


FIGURE 3.4 – Histogramme des fonctions de coût CF2, CF3 et CF4 pour 400 répétitions

comparer le rapport entre la longueur de l'intervalle entre 10^{ème} et 90^{ème} percentiles et la valeur moyenne de la fonction de coût dans une petite zone autour de son minimum. Nous avons choisi une zone à $\pm 5\%$ autour de la valeur de référence (20, 6) de (μ_a, σ_a) , c'est-à-dire une zone de $\mu_a \in [19, 21]$ et $\sigma_a \in [5.7, 6.3]$. Dans cette zone, on choisit 25 points avec $\mu_a \in \{19, 19.5, 20, 20.5, 21\}$ et $\sigma_a \in \{5.7, 5.85, 6, 6.15, 6.3\}$. Les fonctions de coût sont calculées pour ces 25 points avec une taille d'échantillon de 10000. Ensuite, le rapport entre la longueur de l'intervalle entre 10^{ème} et 90^{ème} percentiles correspondant aux différentes tailles d'échantillon et la valeur moyenne de la fonction de coût des 25 points est calculé.

Les résultats sont présentés sur la Figure 3.5. On observe que pour les fonctions de coût CF1, CF3 et CF5, ce rapport converge avec la taille de l'échantillon. Cela montre que la fonction de coût devient plus robuste avec des tailles d'échantillon plus grandes. Pour CF2, le ratio est toujours autour de 0.8, ce qui signifie que la dispersion est importante par rapport à la valeur de la fonction de coût, ce qui rend la fonction coût plus robuste. Le ratio est toujours faible pour les fonctions de coût CF4 et CF6, illustrant que la taille de l'échantillon n'influence pas beaucoup la valeur de la fonction de coût et que la dispersion est relativement faible. Il est donc possible de distinguer les fonctions CF2, CF4 et CF6 qui semblent avoir le même comportement sur la Figure 3.3. On peut en conclure que les fonctions CF4 et CF6 présentent la meilleure robustesse et une convergence très rapide avec la taille de l'échantillon T^{EF} . Dans des cas pratiques, une petite taille d'échantillon peut être utilisée pour économiser du temps de calcul en assurant la robustesse avec ces deux fonctions.

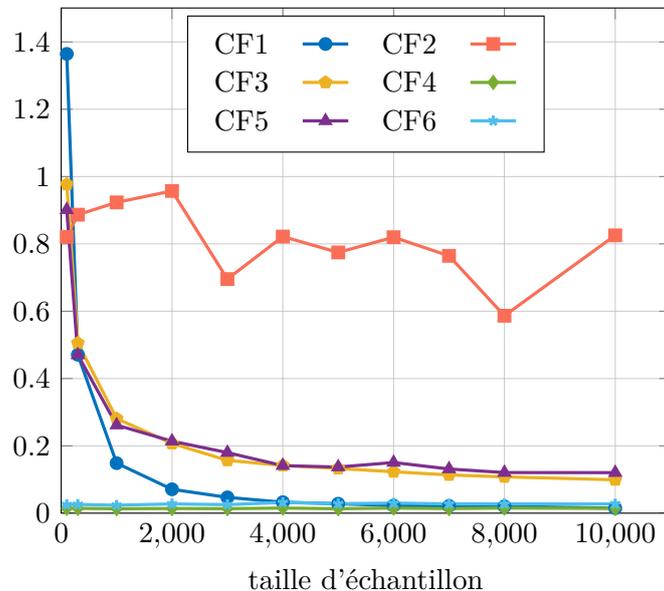


FIGURE 3.5 – Rapport entre la longueur de l'intervalle entre 10^{ème} et 90^{ème} percentiles pour différentes tailles d'échantillon et la valeur moyenne de la fonction de coût dans la zone $\pm 5\%$

3.5.3.3 Identification par algorithme d'optimisation

L'influence de la taille de l'échantillon sur l'identification par l'algorithme d'optimisation est maintenant étudiée. L'algorithme de l'essaim de particules est ici utilisé pour trouver le minimum des fonctions de coût. L'optimisation est testée 400 fois pour chaque taille d'échantillon $n = \{10, 100, 300, 500, 1000, 1500, 2000\}$, puis la valeur moyenne et l'écart-type des paramètres identifiés correspondant à chaque taille d'échantillon sont calculés et présentés sur les Figure 3.6 et Figure 3.7. Les résultats avec la fonction de coût CF2 ne sont pas présentés en raison d'un problème numérique dans l'optimisation. La condition pour appliquer le test de χ^2 est que le nombre d'échantillons dans chaque classe de grandeur pour calculer la statistique de χ^2 doit être supérieur à 5. Cette condition n'est pas assurée d'être satisfaite pour CF2. De plus, l'étude de convergence en fonction de la taille d'échantillon menée précédemment a exclu CF2 des fonctions coût utilisables.

On remarque que toutes les fonctions de coût permettent de trouver les valeurs de référence des paramètres $(\mu_a, \sigma_a) = (20, 6)$ avec une précision croissante avec l'augmentation de la taille de l'échantillon (Figure 3.6 et Figure 3.7). Une convergence du résultat d'identification est donc rencontrée avec l'augmentation de la taille de l'échantillon T^{EF} . Pour la taille d'échantillon égale à 2000, l'erreur relative sur les paramètres μ_a et σ_a identifiées par les fonctions de coût par rapport à la valeur de référence est présentée dans le tableau 3.3. Les paramètres identifiés avec les fonctions coût CF3, CF4, CF5 et CF6 ont une erreur relative plus importante que CF1, cela peut être causé par les biais des tests statistiques [93] [94].

TABLEAU 3.3 – Erreur relative sur les paramètres identifiés pour les fonctions de coût avec une taille d'échantillon égale à 2000

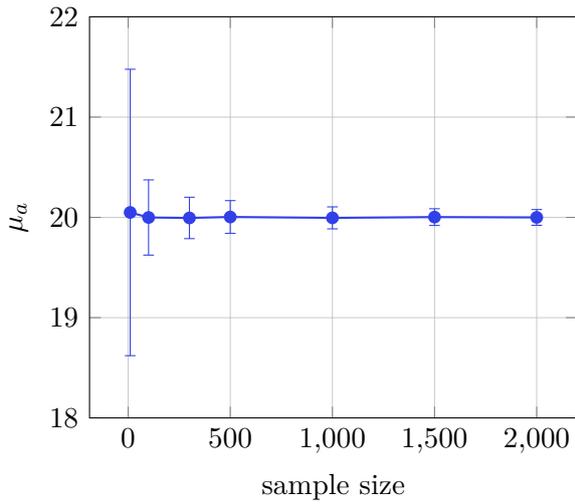
	CF1	CF3	CF4	CF5	CF6
μ_a	-0.0015%	-0.0245%	0.0125%	0.1135%	0.0500%
σ_a	-0.0183%	0.1583%	0.3683%	1.3003%	0.2217%

3.5.4 Performance des fonctions de coût au voisinage du point de référence

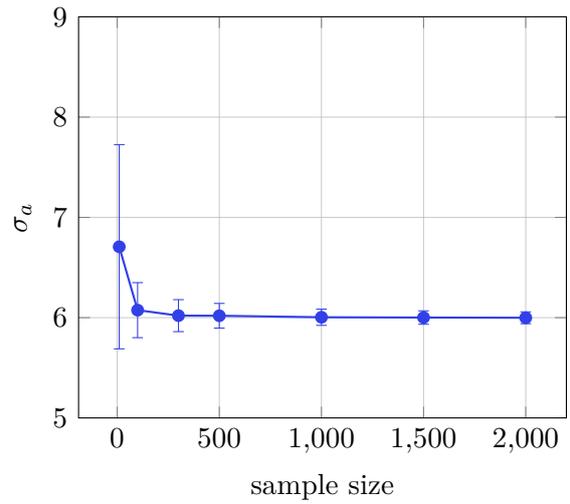
Il est intéressant d'évaluer la précision sur les paramètres identifiés avec l'approche d'identification. Sachant que les fonctions de coût n'ont pas une valeur fixée pour chaque calcul car l'échantillon est généré aléatoirement, il est pertinent de connaître la dispersion des valeurs des fonctions de coût au voisinage du point de référence.

Dans un premier temps, pour $\sigma_a = 6$ fixé et une taille d'échantillon de T^{EF} égale à 2000, les fonctions de coût sont évaluées pour $\mu_a \in [15, 25]$ et 400 fois pour chaque point. La médiane des valeurs des fonctions coût pour chaque paramètre μ_a choisi est calculée, les 10^{ème} et 90^{ème} percentiles fournissent les barres d'erreur sur la Figure 3.8. Il est observé que,

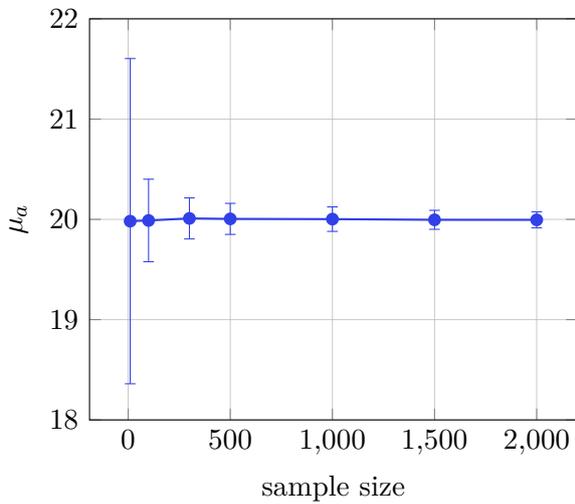
- Pour toutes les fonctions coût, la taille des barres d'erreur est plus petite au point de référence $\mu_a = 20$,



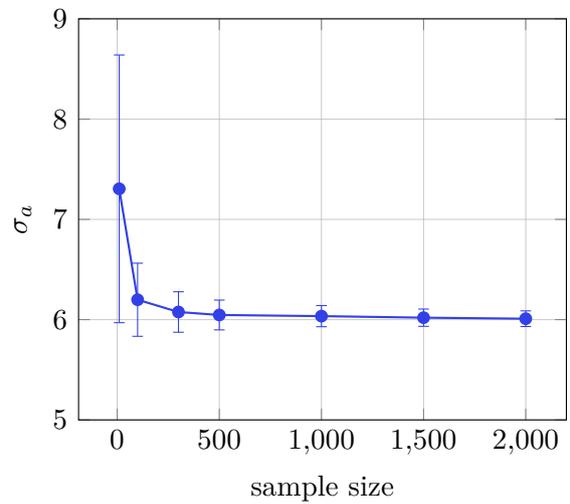
(a) CF1-optimised μ_a



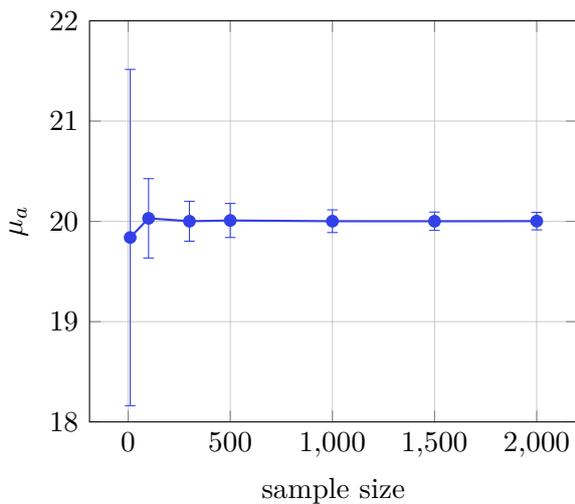
(b) CF1- σ_a optimisé



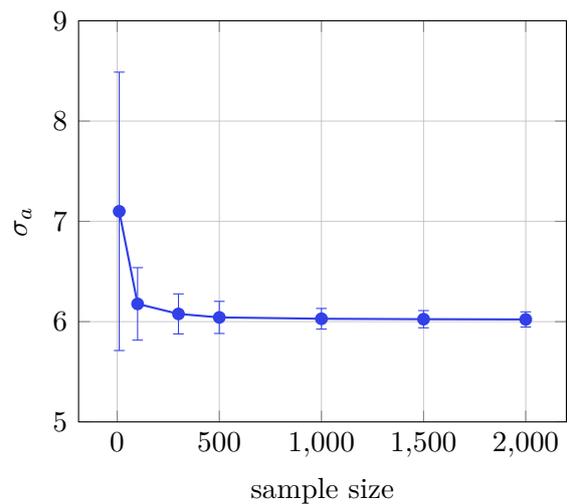
(c) CF3- μ_a optimisé



(d) CF3- σ_a optimisé

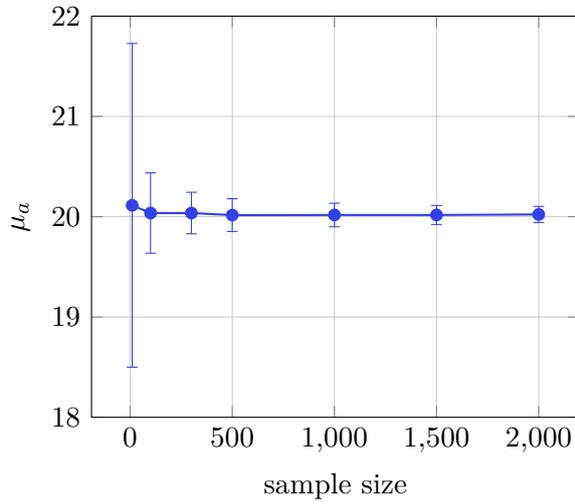


(e) CF4- μ_a optimisé

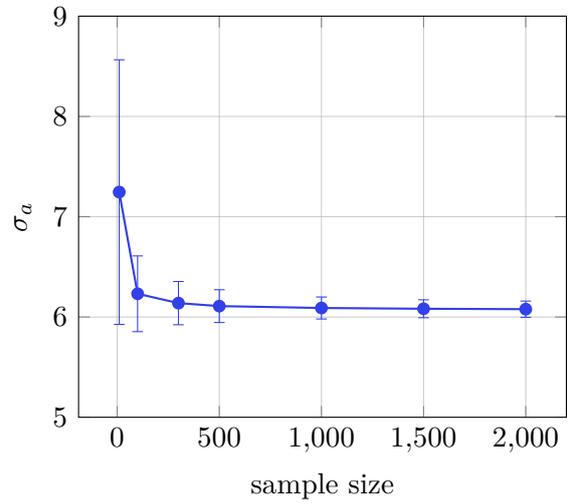


(f) CF4- σ_a optimisé

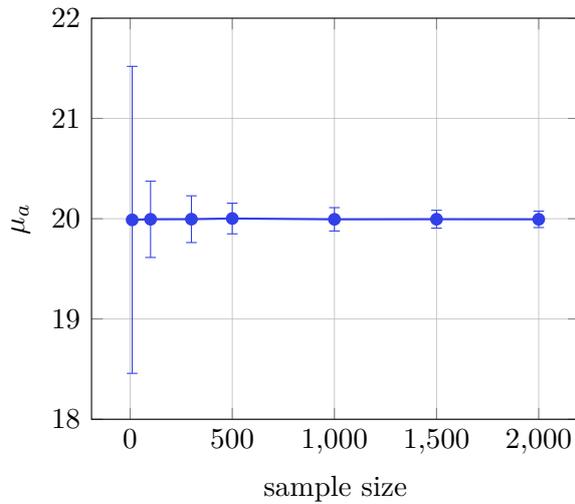
FIGURE 3.6 – Optimisation des paramètres μ_a et σ_a par l'algorithme à essai de particules pour les fonctions coût CF1, CF3 et CF4



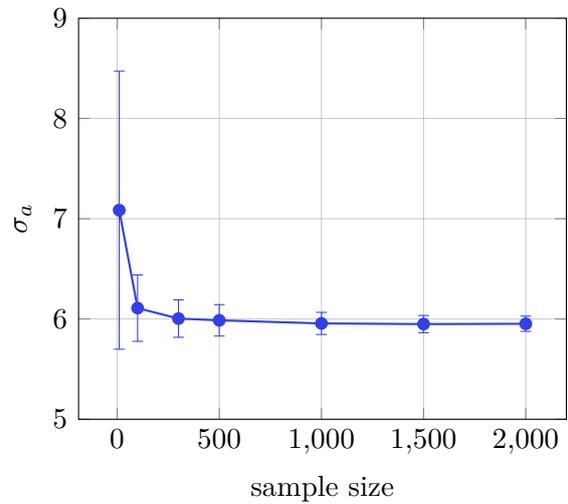
(a) CF5- μ_a optimisé



(b) CF5- σ_a optimisé



(c) CF6- μ_a optimisé



(d) CF6- σ_a optimisé

FIGURE 3.7 – Optimisation des paramètres μ_a et σ_a par l'algorithme à essaim de particules pour les fonctions coût CF5 et CF6

- Pour les fonctions de coût CF1, CF4 et CF6, les barres d'erreur grandissent quand le paramètre μ_a s'éloigne du point de référence,
- Pour les fonctions de coût CF3 et CF5 les barres d'erreur sont plus stables pour différentes valeurs du paramètre,
- La fonction de coût CF2 donne des barres d'erreur plus grandes quand le paramètre μ_a est éloigné de la valeur de référence. Les bornes inférieures des barres d'erreur sont assez proches, ce qui rend difficile l'identification du point de référence.

La même étude a été faite pour $\mu_a = 20$ fixé. Les fonctions coûts sont calculées 400 fois, les médianes, les bornes données par les 10^{ème} et 90^{ème} percentiles sont calculées et présentées dans la Figure 3.9. Contrairement au cas précédent, les fonctions de coût réagissent différemment pour les σ_a plus petits ou plus grands que la valeur de référence.

- Pour la fonction coût CF1, les barres d'erreur sont plus petites pour σ_a petit,
- Les barres d'erreur des fonctions coût CF3 et CF5 restent plutôt constantes sur la plage des valeurs de σ_a , avec une légère décroissance pour les grandes valeurs de σ_a ,
- Pour les fonctions coût CF4 et CF6, les points proches du point de référence $\sigma_a = 6$ ont une barre d'erreur plus étroite, elle augmente quand σ_a s'éloigne du point de référence,
- Pour la fonction de coût CF2, la valeur des fonctions de coût est très sensible au changement de σ_a , les barres d'erreur sont trop grandes pour pouvoir identifier le point de référence.

La même étude est menée sur un voisinage 2D de la valeur de référence, $(\mu_a, \sigma_a) \in [15, 25] \times [2, 10]$. Les fonctions coût sont calculées 400 fois en chaque point d'une grille régulière sur le domaine. Les médianes et barres d'erreur données par les 10^{ème} et 90^{ème} percentiles sont présentées sur la Figure 3.10. Le même comportement est observé en termes de la valeur des fonctions de coût et des barres d'erreur est observé sauf pour la fonction de coût CF2. Pour CF2, il n'y a pas un minimum évident, les comportements des 10^{ème} et 90^{ème} percentiles ne sont pas semblables.

Les comportements pour différentes valeurs de (μ_a, σ_a) ont la même tendance, les surfaces des borne inférieure donnée par le 10^{ème} percentile et borne supérieure donnée par le 90^{ème} percentile sont globalement parallèles.

Une étude de convergence par rapport à la taille d'échantillon est également menée pour un point voisin du point de référence ($\mu_a = 20.5, \sigma_a = 6.5$), car la capacité d'identification des minimums dépend aussi du comportement sur les points voisins. Les fonctions coût sont ainsi évaluées sur le point ($\mu_a = 20.5, \sigma_a = 6.5$) 400 fois pour chaque taille d'échantillon choisi $n = 100, 300, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 10000$, puis la médiane et les 10^{ème} et 90^{ème} percentiles sont calculés et présentés sur la Figure 3.11. Pour CF1, CF3 et CF5, une tendance de convergence est observée en médiane et dispersion par rapport à la taille d'échantillon. En revanche, pour CF4 et CF6, l'estimation ne converge ni en moyenne ni en dispersion. Cela peut expliquer que la valeur de CF4 et CF6 au point de référence ne converge pas avec la croissance de la taille d'échantillon (Figure 3.3), par contre les résultats d'identification par optimisation convergent avec la taille d'échantillon. Parce que la valeur des fonctions coût s'éloigne de celle du point de

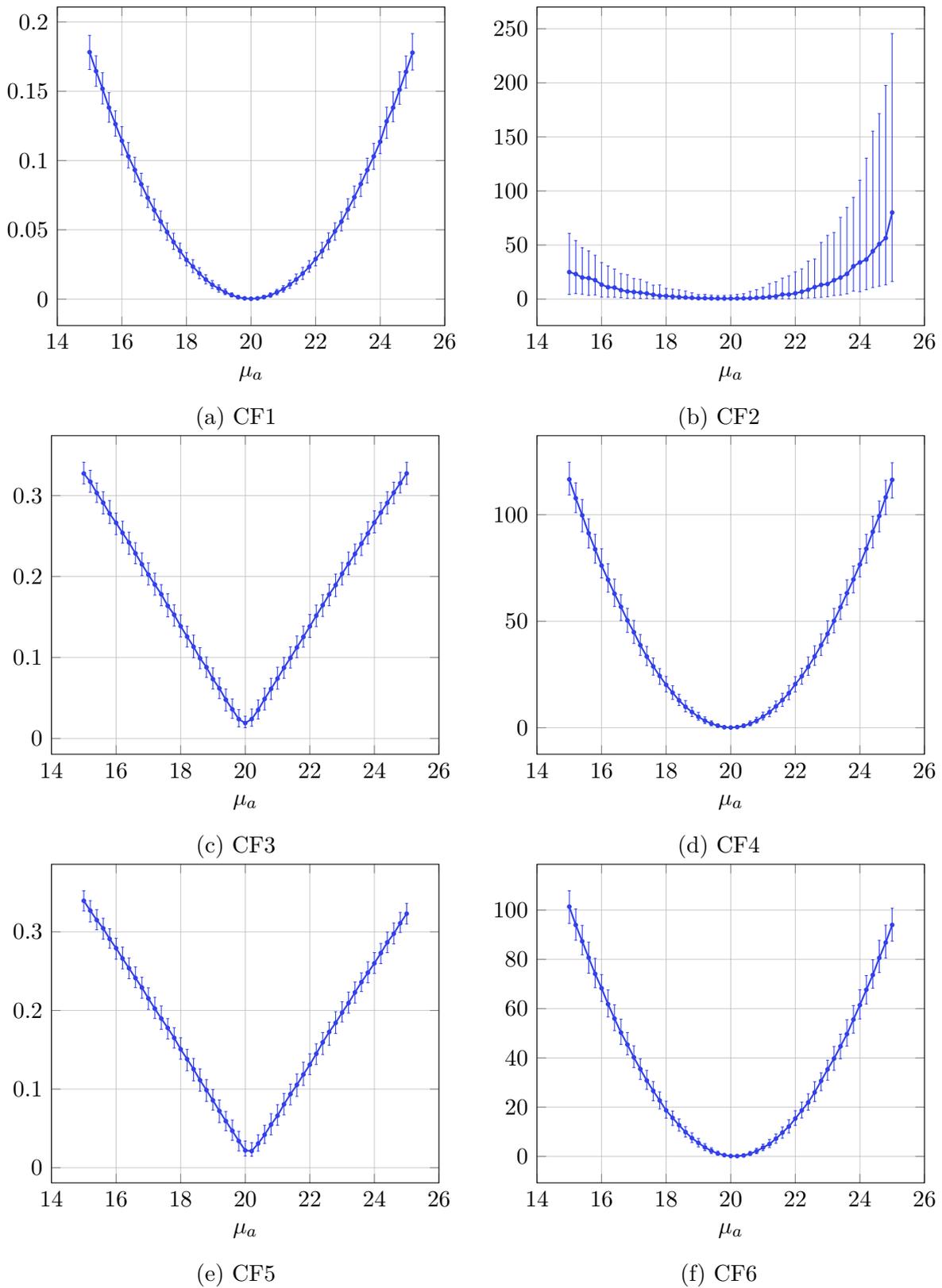


FIGURE 3.8 – Médianes, 10^{ème} et 90^{ème} percentiles des six fonctions de coût avec $\sigma_a = 6$ pour 400 répétitions

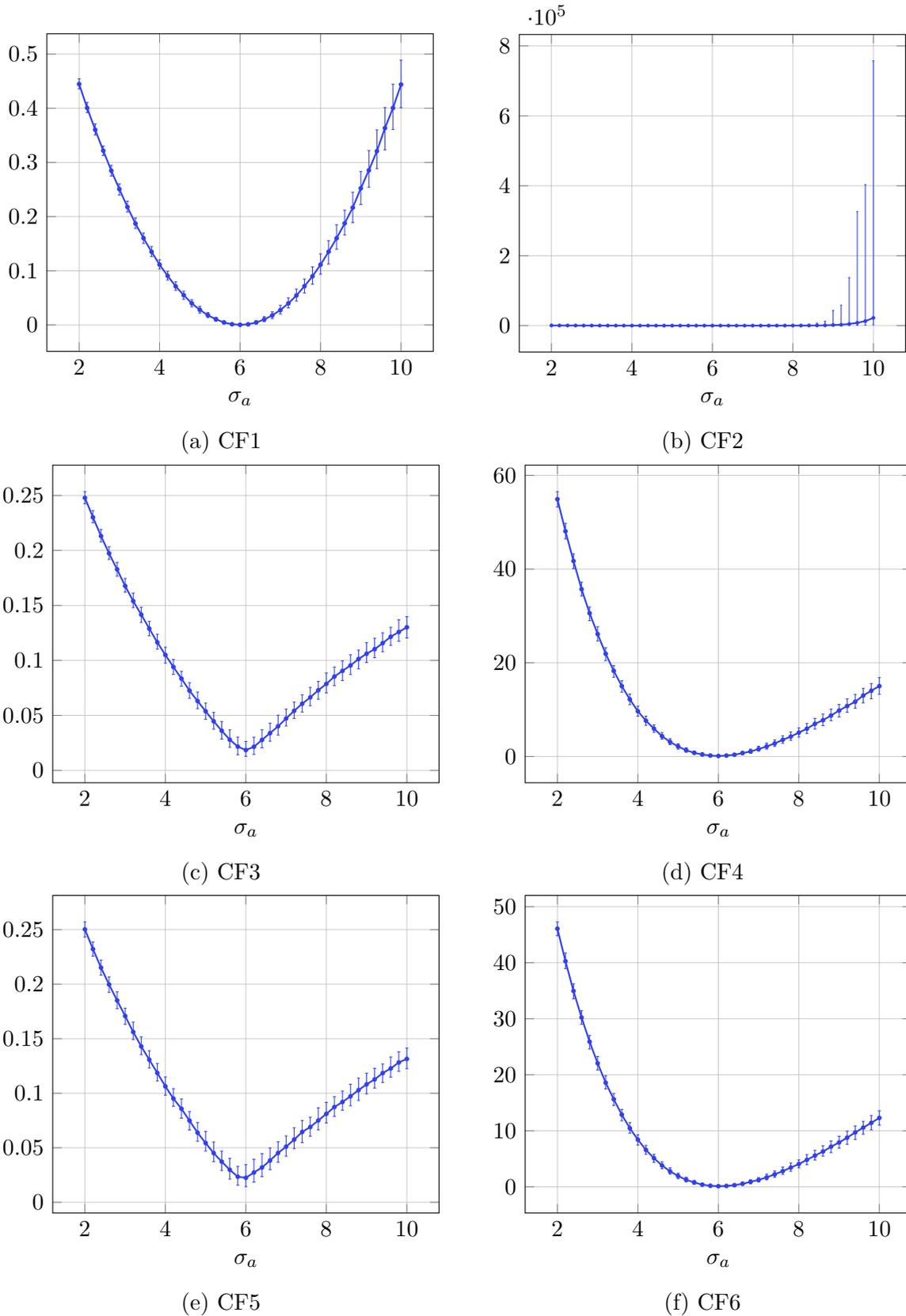


FIGURE 3.9 – Médianes, 10^{ème} et 90^{ème} percentiles des six fonctions de coût avec $\mu_a = 20$ pour 400 répétitions

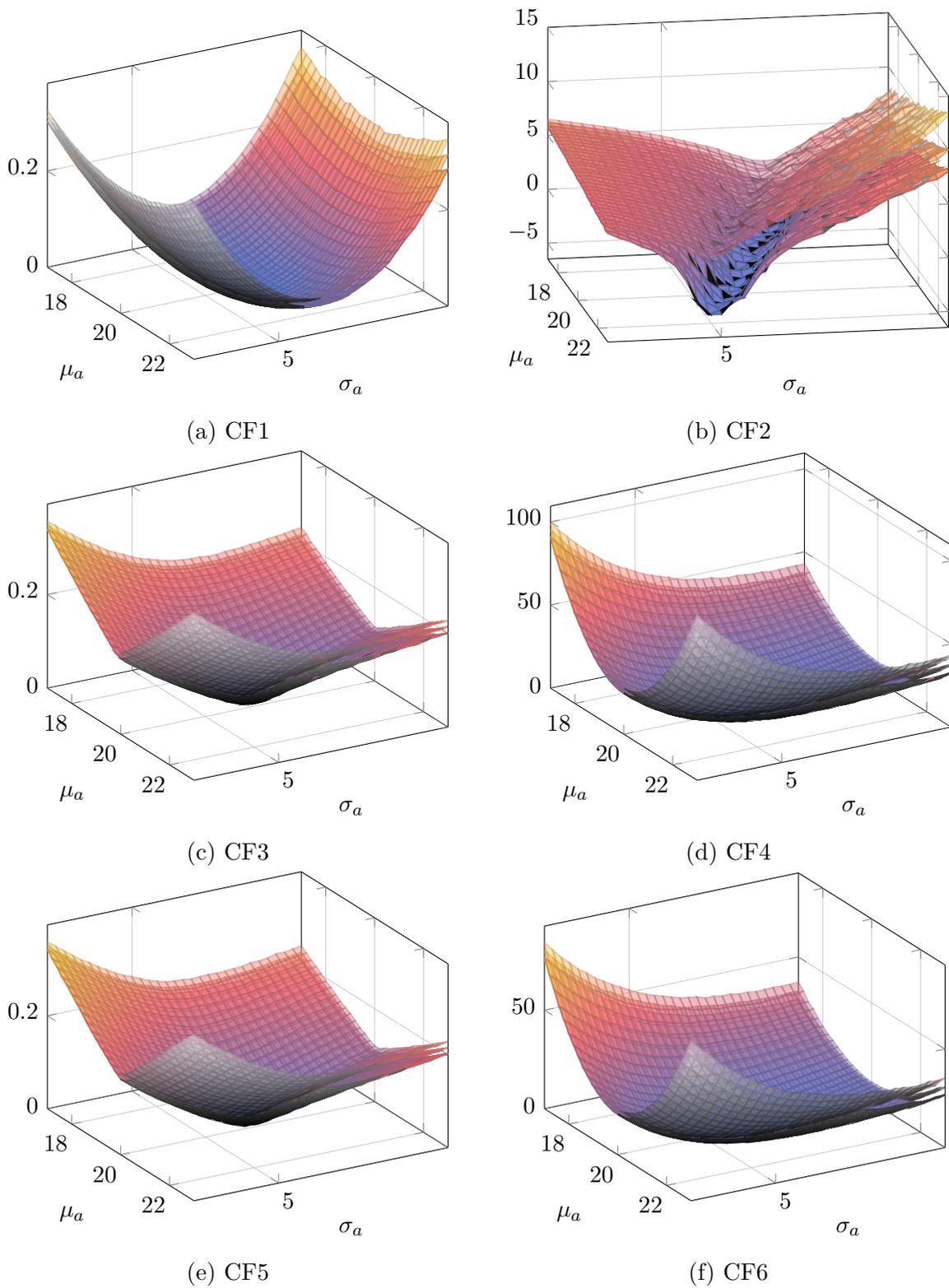


FIGURE 3.10 – La médiane, 10^{ème} et 90^{ème} percentiles des fonctions de coût pour 400 répétitions

référence avec augmentation de la taille d'échantillon. La fonction coût CF2 présente un comportement non régulier.

Selon la Figure 3.6 et la Figure 3.7, la convergence des résultats d'identification est observée pour toutes les fonctions coût avec des tailles d'échantillon similaires de l'ordre de 1000. Pour CF1, CF3 et CF5, la convergence de la valeur de la fonction coût par rapport à la taille d'échantillon est observée pour le point de référence et ses voisins. Pour les fonctions de coût CF4 et CF6, la performance des fonctions coût reste similaire pour différentes tailles d'échantillon selon la Figure 3.3, mais la valeur des fonctions de coût des voisins et les barres d'erreur augmentent avec la taille d'échantillon, cela permet d'identifier plus facilement le minimum par l'algorithme d'optimisation quand la taille d'échantillon augmente.

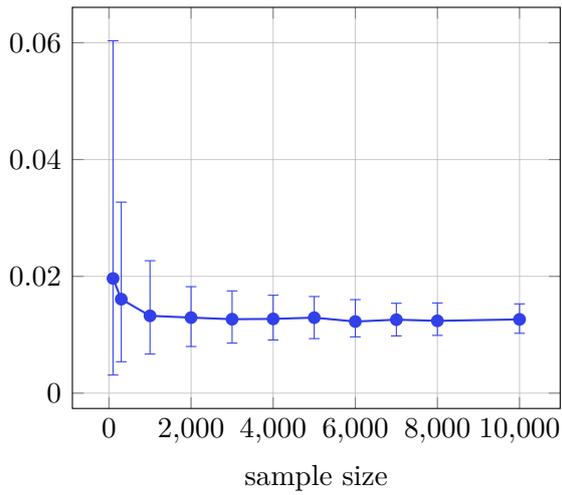
3.6 Cas d'application - identification des trois paramètres de la courbe P-S-N

Dans cette section, nous avons appliqué la méthode d'identification pour trois paramètres de la courbe P-S-N, correspondant aux paramètres (μ_a, σ_a) du paramètre stochastique a et au paramètre b représentant la pente de la courbe de Basquin. La pente b peut être très différente selon les joints de soudure et le matériau. Bien qu'elle soit considérée comme une grandeur déterministe, elle mérite d'être identifiée pour les joints de soudure.

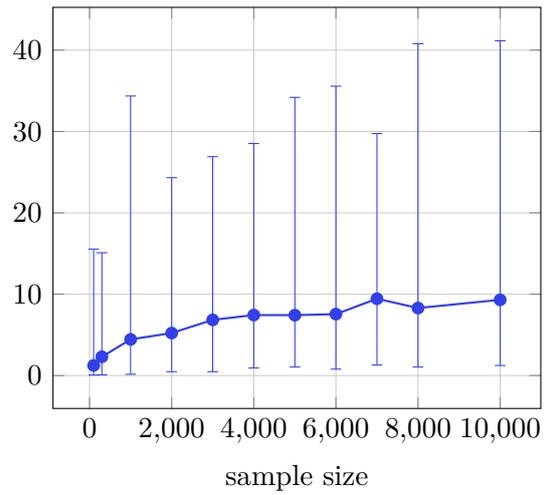
3.6.1 Identification à partir des données d'un seul cordon de soudure

Pour les exemples numériques, les valeurs de référence de ces paramètres sont fixées à $\mu_a = 20$, $\sigma_a = 6$ et $b = 3$. L'identification est étudiée pour les six fonctions de coût et, pour commencer, sur la base des données d'amplitude de la contrainte S et du nombre de cycles correspondant N déduites par la technique Rainflow à partir de la contrainte temporelle éléments finis pour un joint de soudure. L'optimisation est réalisée deux fois pour chaque fonction de coût différente, les résultats du tableau 3.4 montrent que le paramètre σ_a est bien identifié, ce n'est pas le cas pour μ_a et b . Le problème inverse semble être mal posé avec des solutions non uniques. La raison est probablement que le modèle est composé de trop de paramètres par rapport aux données disponibles. Une méthode de régularisation pourrait être utilisée dans ce cas, mais il est difficile de choisir l'information préalable et de garantir sa fiabilité.

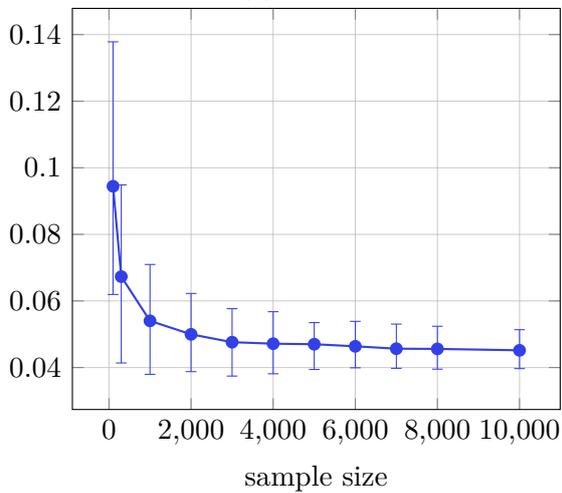
Comme le paramètre σ_a est bien identifié, nous considérons les fonctions de coût dans le plan (μ_a, b) pour la valeur de référence de σ_a . Les valeurs des fonctions de coût CF1 et CF4 sont dessinées dans l'espace des paramètres (μ_a, b) (Figure 3.12), les performances des fonctions de coût CF4, CF5 et CF6 sont très similaires à celles de CF3. Il est observé que le minimum de la fonction coût n'est pas unique. Il est ainsi clair que pour un ensemble de données provenant d'un seul cordon, la solution du problème inverse n'est pas unique.



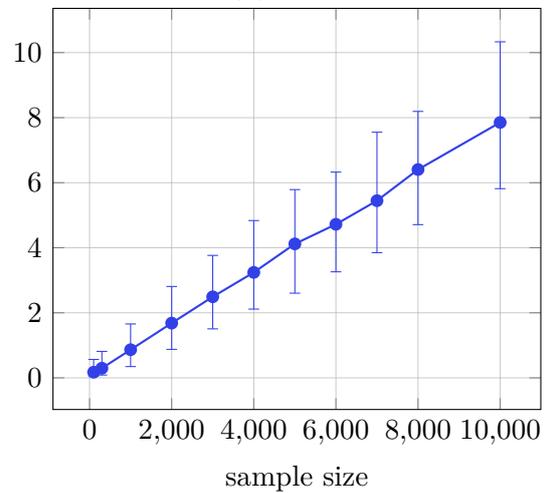
(a) CF1



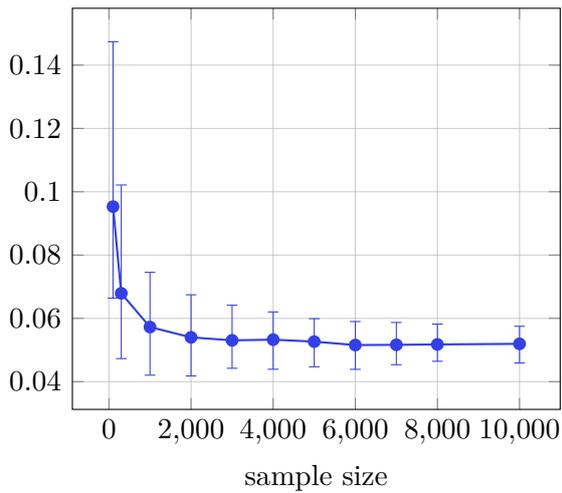
(b) CF2



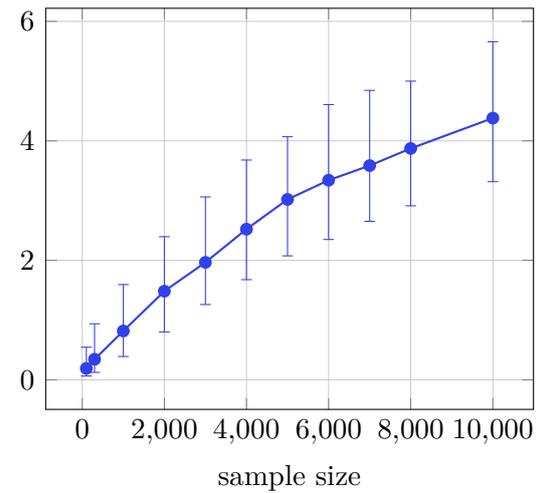
(c) CF3



(d) CF4



(e) CF5

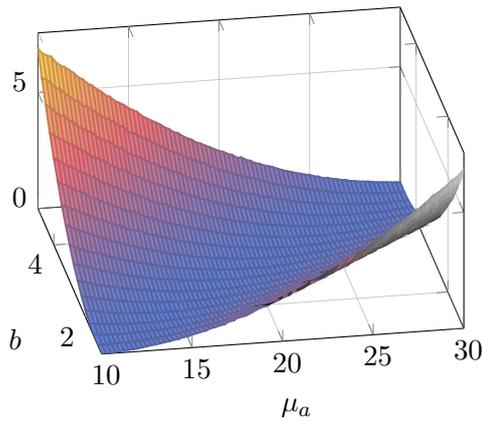


(f) CF6

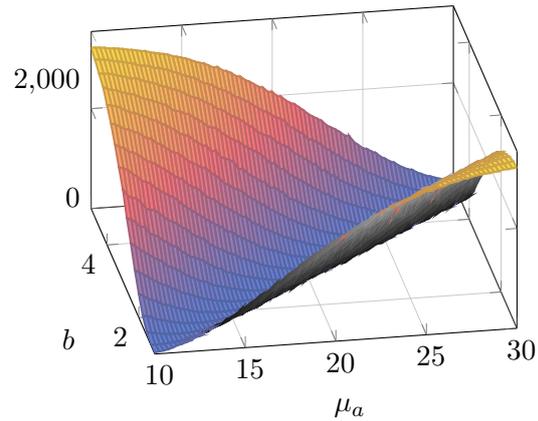
FIGURE 3.11 – Médiane, 10^{ème} et 90^{ème} percentiles des six fonctions de coût pour différentes tailles d'échantillon T^{EF}

TABLEAU 3.4 – Optimisation en deux temps de trois paramètres pour différentes fonctions de coût

	CF1	CF3	CF4	CF5	CF6
μ_a	24.07	43.79	32.59	9.87	13.12
σ_a	6.04	6.09	6.06	5.98	6.02
b	4.75	5.41	3.0	0.99	1.64
μ_a	17.45	28.51	24.92	5.30	9.90
σ_a	5.97	5.96	5.96	5.95	6.04
b	3.54	4.67	3.94	0.0005	0.98



(a) CF1



(b) CF4

FIGURE 3.12 – Fonctions de coût CF1 et CF4 dans l'espace des paramètres pour σ_a fixé

3.6.2 Identification à partir des données de plusieurs cordons de soudure

Considérant que davantage de données peuvent enrichir le modèle stochastique, nous menons l'identification à partir des données de quatre joints de soudure situés dans des zones différentes de la structure. Les données amplitude de contraintes S et nombre de cycles N correspondantes (données des zones 1 à 4) calculées à partir de la contrainte temporelle des quatre joints de soudure, et des durée de vie expérimentales fabriquées avec les paramètres de référence sont utilisées. Le problème d'optimisation est traité en utilisant la fonction de coût multi-sites présentée dans la Section 3.4.2.3. Finalement, une vallée de minimums des fonctions de coût est encore observée sur une ligne droite dans la Figure 3.13. Cependant un minimum global difficile à observer est trouvé proche de la valeur de référence de μ_a et b sur cette ligne. Cela signifie qu'il est possible de réaliser l'identification en augmentant la qualité des données, malgré une faible sensibilité le long de cette ligne.

Pour mieux comprendre la signification physique de la vallée minimale, la courbe P-S-N pour des jeux de paramètres le long de la vallée et les ensembles des données sont présentées (Figure 3.14). Sur la figure, la dispersion des courbes S-N est représentée

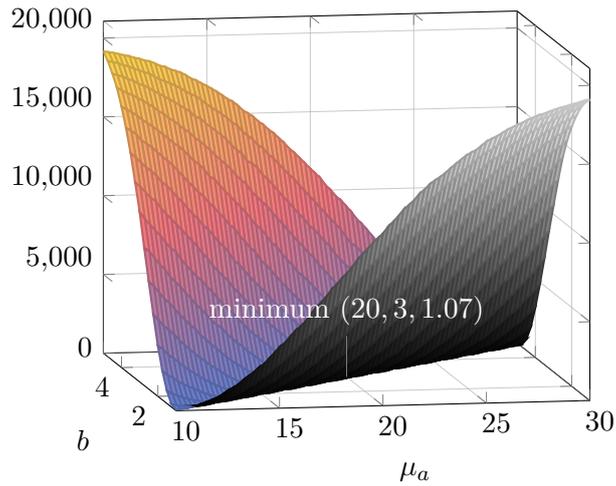


FIGURE 3.13 – Fonction de coût CF4 dans l'espace paramétrique pour σ_a fixé avec multi joints de soudure

par σ_a . Il est montré que les données sont encerclées par le groupe de courbes P-S-N. Les dommages calculés et leur dispersion peuvent être très similaires, ainsi la fonction de coût est très plate sur la vallée. Il faudrait faire attention que S et N sur la Figure 3.14 sont sur l'échelle logarithme naturel (base e). Les courbes de Wöhler sont souvent décrites en échelle de logarithme décimal (base 10) en pratique. Pour mieux comprendre la signification physique des courbes obtenues, la Figure 3.15 est en échelle de logarithme décimal avec les même données dans la Figure 3.14.

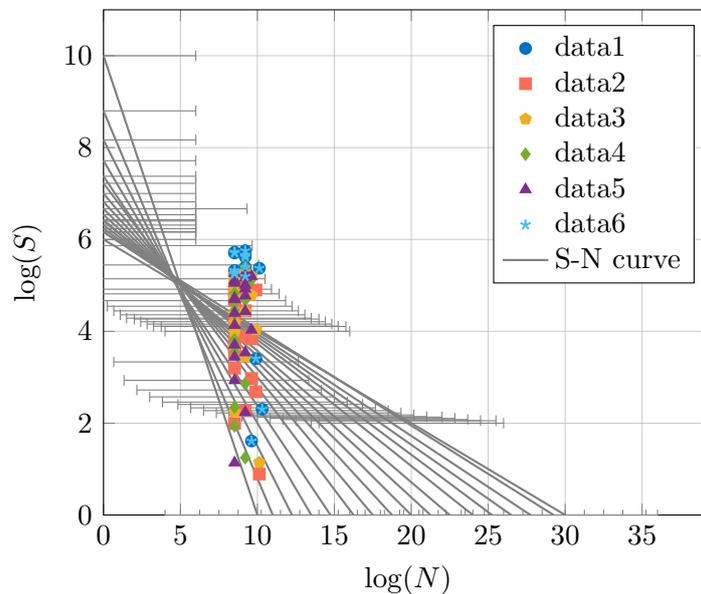


FIGURE 3.14 – Groupe de la courbe S-N optimisée

On observe que, pour différents jeux de données chacun correspondant à une position de joint soudé, l'emplacement de la vallée minimale est légèrement différent, comme le montre sur la Figure 3.16(b). Deux autres ensembles de données de S_i et n_i (données 5 et données 6), présentant une différence relativement importante, sont choisis pour effec-

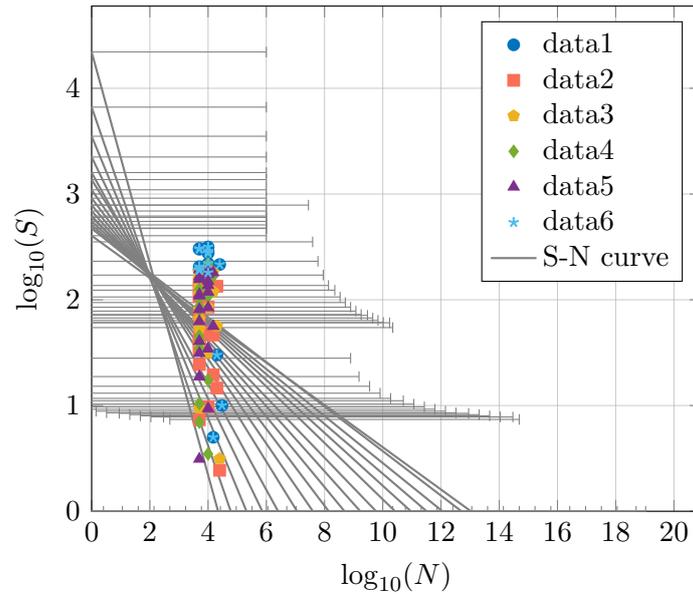
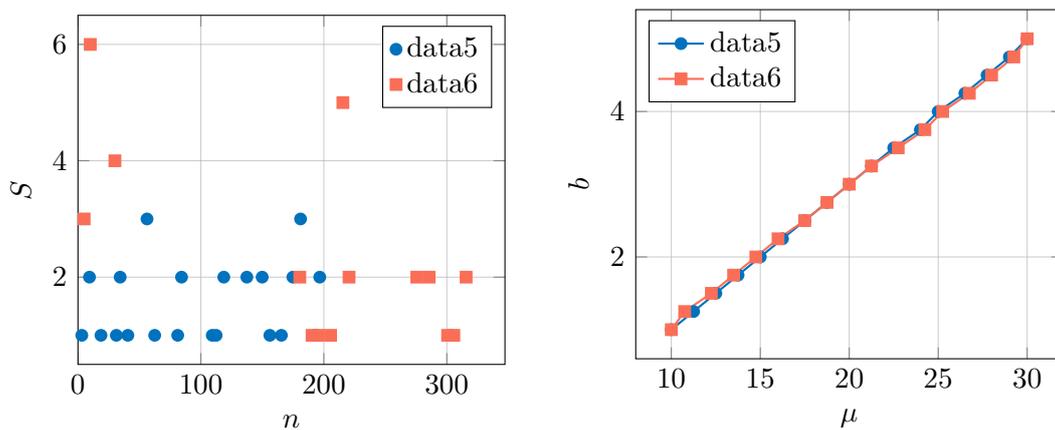


FIGURE 3.15 – Groupe de la courbe S-N optimisée en échelle de logarithme décimal

tuer le test, les données sont illustrées sur la Figure 3.16(a). L'identification est effectuée séparément avec chaque ensemble de données par la Figure 3.16(b), ce qui conduit dans chaque cas à une ligne de solutions non uniques. L'intersection des deux lignes correspond au minimum global de la fonction de coût basée sur les deux ensembles de données réunis. L'utilisation des deux ensembles de données conduit à une solution unique. Pourtant, les lignes minimales correspondant à chaque ensemble de données de la fonction coût multi-sites ne sont pas très différentes. Cela peut conduire à une vallée prononcée de la fonction coût multi-sites qui peut rendre la minimisation difficile.



(a) Présentation des données

(b) Localisation des minimums pour CF3 avec deux ensembles des données

FIGURE 3.16 – Fonctions de coût dans l'espace des paramètres avec σ_a fixé

Il semble possible de réaliser l'identification avec davantage de données, ainsi l'optimisation des fonctions de coût multi-sites construites avec l'ensemble de données est effectuée, et les résultats sont présentés dans le tableau 3.5. Avec quatre ensembles de données, l'optimisation directe avec trois paramètres donne un résultat satisfaisant avec

les fonctions de coût CF4 et CF6. Cependant, s'il y a moins de données disponibles, il est plus difficile pour l'optimisation directe avec trois paramètres de déterminer la solution unique de μ_a et b .

TABLEAU 3.5 – Optimisation des trois paramètres avec différentes fonctions de coût

	CF1	CF3	CF4	CF5	CF6
μ_a	18.63	21.14	20.41	18.58	20.46
σ_a	6.02	5.99	6.04	6.07	6.02
b	2.74	3.21	3.08	2.72	3.08

3.6.3 Stratégie d'optimisation multi-niveaux adaptée

Une stratégie est proposée pour effectuer l'optimisation de la fonction de coût en deux étapes avec moins de données. Tout d'abord, les trois paramètres sont déterminés en minimisant les fonctions de coût, le paramètre σ_a optimisé est conservé. Ensuite, avec la valeur optimisée de σ_a , l'optimisation est effectuée pour les paramètres μ_a et b . Cette stratégie est appliquée avec les données 5 et 6, les résultats sont présentés dans le tableau 3.6. Il est montré que la stratégie fonctionne bien pour les fonctions de coût CF1, CF4 et CF6 avec des ensembles de données de seulement deux joints de soudure, mais elle n'est pas suffisante pour CF3 et CF5. Ces résultats sont conformes à l'analyse de la Figure 3.5. Avec la même quantité de données, lorsque la taille de l'échantillon de T^{EF} est de 10000, les fonctions de coût CF1, CF4 et CF6 sont plus robustes que CF3 et CF5. Cependant, il est attendu que lorsque le nombre de joints de soudure augmente, les fonctions de coût CF3 et CF5 pourraient fonctionner pour l'identification.

TABLEAU 3.6 – Optimisation des 3 paramètres en deux étapes pour les différentes fonctions coût

		CF1	CF3	CF4	CF5	CF6
Étape I	μ_a	18.43	17.66	17.77	21.20	17.09
	σ_a	6.01	6.12	6.04	5.94	5.98
	b	2.70	2.56	2.59	3.23	2.45
Étape II	μ_a	20.29	18.89	20.40	21.56	20.92
	b	3.06	2.79	3.08	3.29	3.17

Une stratégie d'optimisation à plusieurs niveaux pour identifier les paramètres μ_a, σ_a pour a et la pente déterministe b est proposée et testée. L'optimisation directe pour trois paramètres est possible lorsqu'il y a suffisamment de données et que les données couvrent différentes plages d'amplitudes de contraintes. Une stratégie d'optimisation en deux étapes est recommandée nécessitant moins de données. Les fonctions de coût CF1, CF4 et CF6 semblent être plus robustes vis-à-vis de la quantité de données.

L'optimisation de trois paramètres permet d'identifier non seulement les paramètres stochastiques μ_a, σ_a du modèle de courbe P-S-N, mais aussi le paramètre déterministe b . La pente b influence beaucoup le calcul de l'endommagement et de la durée de vie, ce

paramètre est à identifier. Ces paramètres identifiés doivent permettre de conduire à une simulation numérique plus proche des résultats des essais expérimentaux.

3.6.4 Identification des trois paramètres pour une autre référence

Les méthodologies d'identification sont ici appliquées pour une autre référence des paramètres, soient $\mu_a = 30, \sigma_a = 0.5$ et $b = 3$. Ce jeu de paramètres est proche de celui des applications Renault. Cette étude permet aussi d'observer le comportement des méthodologies d'identification sur différents jeux de données.

Pour $\sigma_a = 0.5$ fixé, les fonctions de coût sont évaluées dans l'espace des paramètres (μ_a, b) . Le comportement est similaire à celui pour la référence $(20, 6, 3)$, une vallée de minimums des fonctions est observée. L'étude est ensuite menée pour les fonctions coût multi-sites. Pour 2 jeux de données, avec σ_a fixé, un minimum global est trouvé au point de référence $\mu_a = 30, b = 3$ (Figure 3.17), il est davantage marqué que pour la référence précédente. En revanche, la fonction de coût CF2 induit un problème de calcul sur une moitié de l'espace.

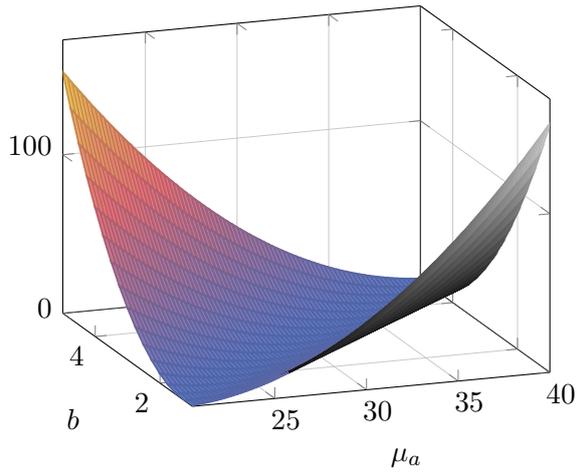
Les courbes S-N correspondant à la vallée des minimums sont représentées ainsi que les couples d'amplitude de contrainte et de nombre de cycle associés (Figure 3.18). Avec la référence $\mu_a = 30, \sigma_a = 0.5$ et $b = 3$, les courbes S-N sont plus raisonnables par rapport aux positions des jeux de données. Ces données se situent en dessous des courbes, cela signifie que chaque amplitude de contrainte mène un dommage bien inférieur à 1, ce qui est réaliste.

Les résultats avec la seconde référence montrent que les méthodologies d'identification des trois paramètres de la courbe de Wöhler probabiliste fonctionnent pour différents jeux de valeurs. Pour le deuxième jeu de valeurs de référence, proche de celui de l'industriel, les méthodologies présentent une meilleure performance. Cela signifie que les méthodologies devraient donner des résultats satisfaisants avec les données expérimentales de Renault.

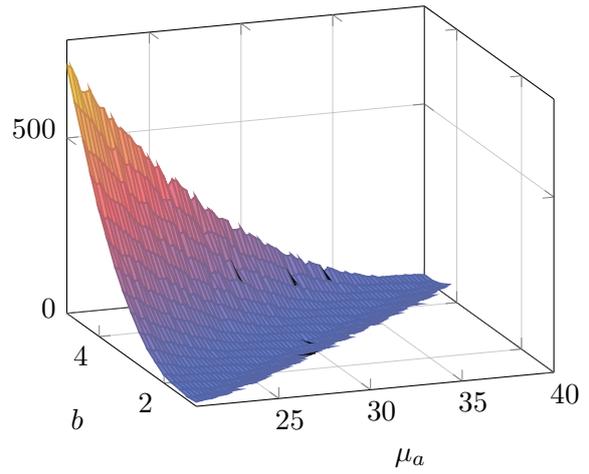
3.7 Précision des méthodologies d'identification

Une étude de simulation est ici menée sur la précision des méthodologies d'identification global tel que présenté sur la Figure 1.17. Dans notre processus d'identification, l'erreur peut venir tout d'abord des données groupées, car il y a un manque d'information par rapport aux données non-censurées. Les estimateurs des paramètres de la loi Log-normale obtenue des données groupées ont une incertitude, qui peut être évaluée par l'information de Fisher pour l'estimation du maximum de vraisemblance et la loi *a posteriori* pour l'estimation bayésienne (voir Chapitre 2). L'erreur peut également provenir de la méthodologie car les fonctions de coût sont construites avec un échantillon par tirage, ce qui mène aux incertitudes étudiées dans les sections 3.5.3, 3.5.3.3 et 3.5.4. Dans cette section, une étude de simulation est proposée pour évaluer une notion d'erreur sur les méthodologies.

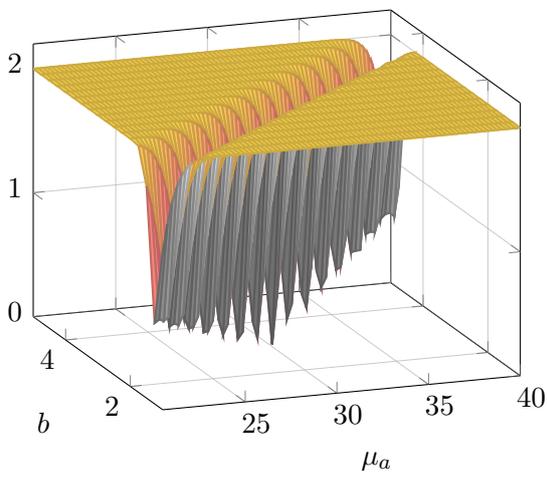
L'objectif de ce chapitre est d'évaluer la variabilité des paramètres μ_a et σ_a , identifiés à partir d'un jeu de données de durée de vie, groupées et construites numériquement



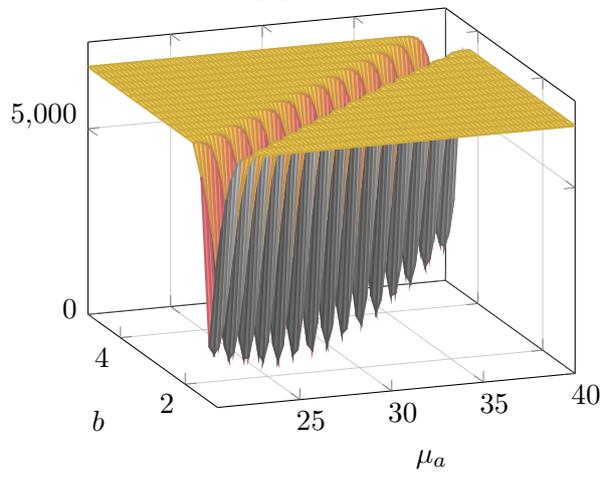
(a) CF1



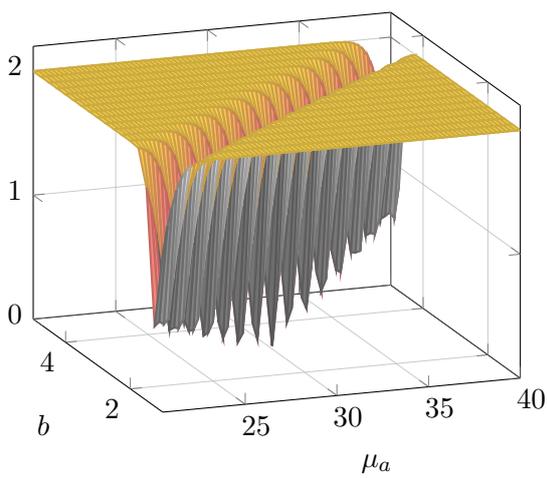
(b) CF2



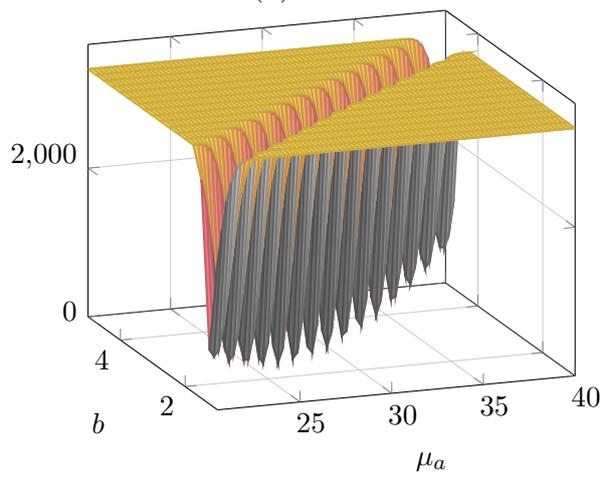
(c) CF3



(d) CF4



(e) CF5



(f) CF6

FIGURE 3.17 – Valeur des six fonctions de coût dans l'espace des paramètres (μ_a, σ_a) avec la seconde référence

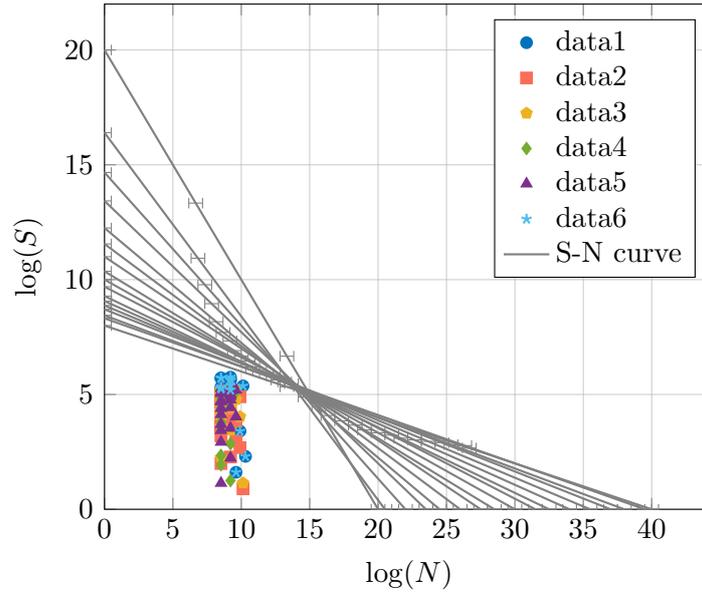


FIGURE 3.18 – Courbes S-N optimisées et données de 1 à 6

avec les méthodologies d'identification. Afin de maîtriser le temps de calcul de l'étude, un réseau de fonctions à base radiale (RBF) est utilisé comme surface de réponse en remplaçant des fonctions coût.

Tout d'abord, les durées de vie sont générées à partir des séquences d'amplitudes de contraintes S_i et le nombre n_i de cycles correspondant comptés par la technique Rainflow, avec une courbe de Wöhler probabiliste de référence.

Pour simuler des données représentatives des données Renault, la valeur de référence de μ_a de Renault est prise et la valeur de référence de σ_a est choisie comme une valeur proche du σ_a identifiée au Chapitre 4 sur les données réelles. Pour des raisons de confidentialité, le paramètre μ_a de référence n'est pas présenté, les valeurs identifiées seront présentées en terme d'erreur relative par rapport à la valeur de référence, et $\sigma_a = 0.35$. Les tirages n_r du paramètre a sont effectués pour générer différentes courbes de Wöhler, puis les échantillons de durée de vie sont calculés. Le nombre de tirages du paramètre a et la taille d'échantillon sont de 40, ce qui est représentatif du nombre d'observations sur l'essai synthétique industriel.

Pour simuler le fait que les durées de vie réelles sont groupées, les échantillons de durée de vie sont regroupés dans les intervalles de temps normalisés entre les bornes : $\{0; 50\%; 100\%; 150\%; 200\%; \infty\}$ pour générer les données groupées. La méthode d'estimation du maximum de vraisemblance est utilisée pour estimer la loi Log-normale des durées de vie expérimentales. En même temps, les lois asymptotiques des estimateurs $\hat{\mu}_t$ et $\hat{\sigma}_t$ sont obtenues. De plus, dans le Chapitre 2, il a été montré que l'écart entre l'estimateur de maximum de vraisemblance et la valeur vraie du paramètre suit asymptotiquement une loi Normale centrée sur la vraie valeur avec une variance décrite par l'information de Fisher (Équation (2.14) dans Chapitre 2). L'incertitude liée à l'estimateur de maximum de vraisemblance et la dispersion obtenue sur plusieurs jeux de tirages de durées de vie groupée seront comparées. Les méthodologies d'identification sont appliquées à partir de la loi estimée des durées de vie expérimentales paramétrée par μ_t et σ_t pour identifier μ_a

et σ_a . Un réseau de fonctions à base radiale (RBF) est construit avec les entrées μ_t et σ_t et les sorties μ_a et σ_a pour pouvoir accélérer les processus d'identification. La mise en place de cette étude sur la précision d'identification est divisée en 2 étapes, premièrement la génération des données de durée de vie, et deuxièmement la construction du RBF. Les résultats seront ensuite analysés.

3.7.1 Génération des échantillons de μ_t et σ_t

L'échantillon de μ_t, σ_t peut être obtenu par deux approches :

- faire n_r tirages de durées de vie groupées puis faire l'estimation par le maximum de vraisemblance ;
- générer l'échantillon avec 1 tirage de durées de vie par la loi asymptotique de l'estimateur de maximum de vraisemblance.

Les deux approches ne représentent pas la même chose, soit l'incertitude liée au tirage sur la durée de vie et le fait de les grouper, soit l'incertitude sur la méthode du maximum de vraisemblance.

3.7.1.1 Tirage des durées de vie groupées et estimation

A partir du tirage du paramètre a selon la loi de référence $a \sim (\mu_a^{ref}, \sigma_a^{ref})$ de taille 40, la durée de vie exacte non-censurée est calculée en utilisant les contraintes et nombre de cycles pour un des cordons de soudure, par une analyse Rainflow du calcul EF. Les durées de vie exactes sont placées dans les intervalles de temps normalisés entre les instants 0; 50%; 100%; 150%; 200%; ∞ . L'estimation du maximum de vraisemblance est réalisé pour obtenir μ_t et σ_t à partir des données groupées. Le processus est répété n_r fois, il fournit un échantillon de μ_t et σ_t de taille n_r .

3.7.1.2 Tirage de μ_t et σ_t à partir de l'estimateur du maximum de vraisemblance

Comme ci-dessus μ_t et σ_t sont estimés à partir d'un tirage de durées de vie groupées qui est ici unique, l'écart entre l'estimateur et la vraie valeur suit une loi Normale asymptotique comme vu au Chapitre 2 dans l'équation (2.14), et la distribution de la vraie valeur du paramètre par rapport à l'estimateur est connue. Un tirage de μ_t et σ_t est réalisé selon cette loi. Pour faire le tirage, deux méthodes d'échantillonnage sont comparées, celle par hypercube latin et celle par la séquence de Halton.

3.7.1.2.1 Méthode d'échantillonnage

Grâce à la méthode de Box-Muller, la génération d'un échantillon d'une loi continue peut se faire à partir de la génération d'un échantillon de la loi Uniforme $\mathcal{U}[0, 1]$. La variable aléatoire par transformation de la variable aléatoire par sa fonction de répartition suit une loi uniforme $\mathcal{U}[0, 1]$. L'échantillon de la variable aléatoire peut être obtenu par la méthode de la fonction inverse à partir d'un échantillon uniforme. La génération d'une variable aléatoire (*v.a.*) est réalisée sur ordinateur par une procédure déterministe, par exemple, les suites congruentes et la séquence de Halton. Donc, ces *v.a.* sont considérées

pseudo-aléatoire. Dans cette section, les échantillonnages aléatoires basés sur la séquence de Halton et la méthode d'hypercube latin sont testés.

La séquence de Halton se fait par l'expansion des entiers naturels dans une base, e.g. 2 ou 10, puis la fonction inverse radicale est appliquée pour obtenir un nombre aléatoire entre 0 et 1 [95]. Si un échantillon multidimensionnel est demandé, la base est différente pour chaque dimension. Dans la fonction MATLAB, on peut choisir de sauter les premières itérations, ou prendre le chiffre généré toutes les k étapes, $k \in \mathbb{N}^*$. Cela évite de générer la même séquence à chaque fois.

La méthode d'échantillonnage par hypercube latin, initialement proposée par W.J Conover, est un compromis entre l'échantillonnage aléatoire et l'échantillonnage stratifié [96]. L'idée principale est de diviser l'espace d'échantillonnage en intervalles disjoints équiprobables. Une valeur est sélectionnée dans chaque intervalle. Pour échantillonner des paramètres multidimensionnels indépendants, le traitement précédent s'applique pour chaque paramètre et se met en paire aléatoire entre les dimensions. De cette façon, l'échantillon sera bien distribué uniformément dans l'espace des paramètres.

Voici un exemple d'échantillons générés par l'échantillonnage aléatoire basé sur la séquence de Halton et la méthode hypercube latin. Deux variables indépendantes suivant la loi normale sont générées, soit $X \sim \mathcal{N}(0, 1)$ et $Y \sim \mathcal{N}(1, 2)$.

Tout d'abord, un échantillon uniforme en dimensions deux est généré pour chacune des méthodes. Pour la séquence de Halton, on a décidé de sauter les 100 premiers pas et de prendre une valeur tous les 100 pas. Les résultats sont présentés dans la Figure 3.19.

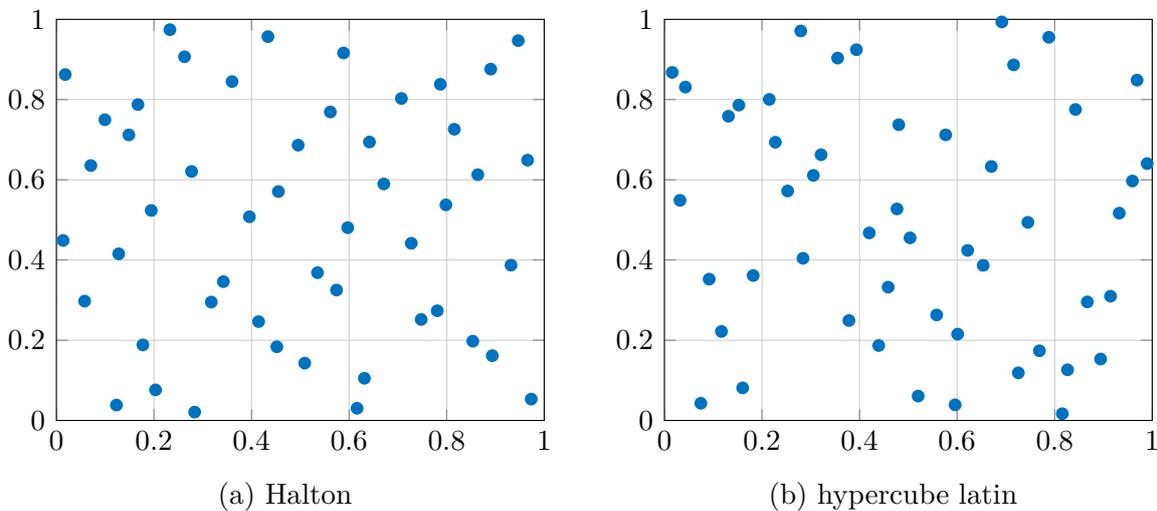


FIGURE 3.19 – Échantillon uniforme généré par la séquence de Halton et par la méthode d'hypercube latin

Puis, l'échantillon des deux variables aléatoires est construit à partir de l'échantillon uniforme à l'aide de l'inverse de la fonction de répartition. L'échantillon uniforme selon l'axe horizontal est pris pour la variable X et selon l'axe vertical pour la variable Y dans la Figure 3.19. L'échantillon de X et Y est représenté dans Figure 3.20. Il est observé que l'échantillon par les deux approches couvre bien l'espace des deux variables.

Selon l'exemple étudié, les deux méthodes donnent des résultats similaires. La méthode de l'hypercube latin est choisie pour échantillonner μ_t et σ_t , car cette méthode

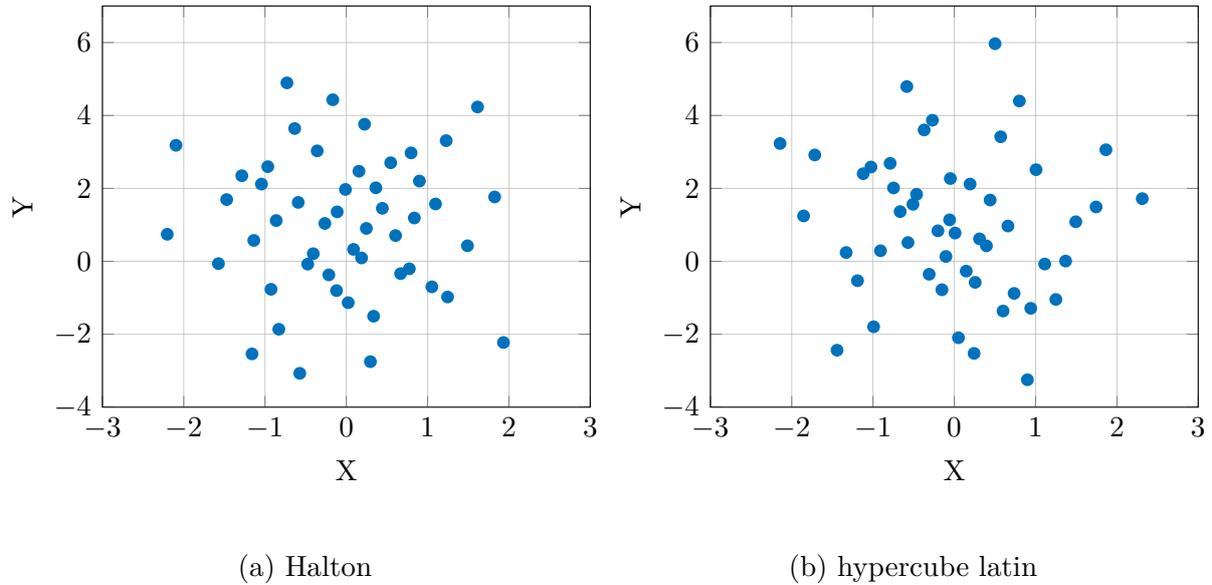


FIGURE 3.20 – Échantillon Normal généré par la séquence de Halton et la méthode d'hypercube latin

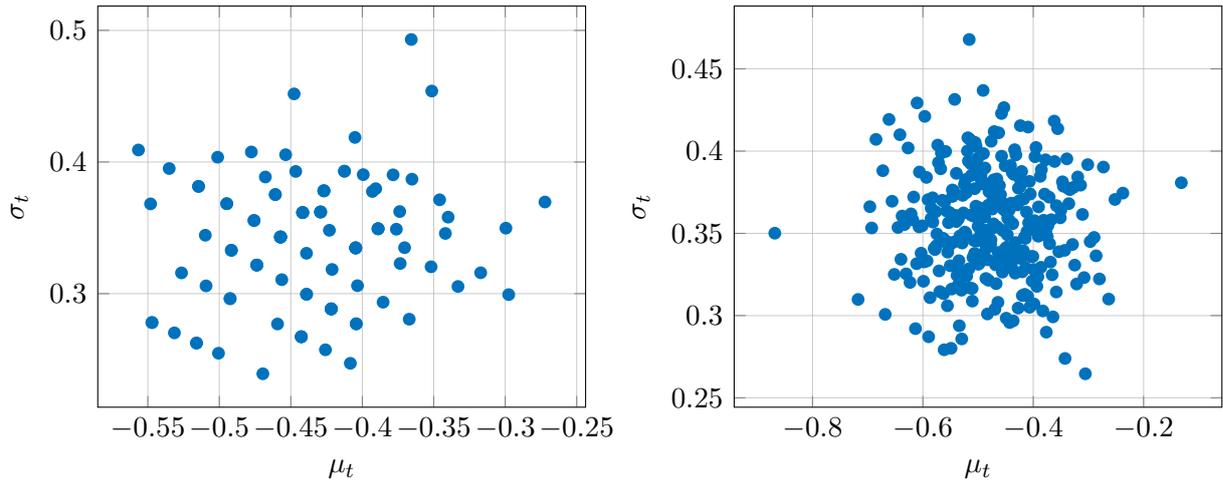
permet de générer plus facilement un échantillon aléatoirement dans l'espace.

3.7.1.3 Comparaison des échantillons obtenus à partir des deux types de variabilité

Les échantillons des paramètres μ_t et σ_t issus des deux types de variabilité, soit par tirage des durées de vie groupées, soit par l'estimateur du maximum de vraisemblance, sont comparés avec une taille égale à 100 dans la Figure 3.21. Pour la première méthode (Figure 3.21(a)), 100 tirages de durées de vie de taille 40 sont réalisés, et pour la deuxième méthode (Figure 3.21(b)), l'échantillonnage par hypercube latin de la loi asymptotique de l'erreur d'estimation est réalisé avec une taille de 100.

Dans la Figure 3.21a, certains points sont superposés, du fait du groupement des durées de vie, cela augmente la possibilité d'avoir le même nombre de données dans chaque intervalle, et donc le même résultat d'estimation. Ainsi à partir d'une même loi de distribution pour les données exactes ($T \sim \mathcal{LN}(-0.4296, 0.3500)$ μ_t, σ_t calculé à partir de μ_a, σ_a avec l'approche analytique la Section 3.3), le fait de grouper génère une perte d'information. L'estimation fournit un paramètre μ_t compris entre -0.6 et -0.25 et le paramètre σ_t compris entre 0.2 et 0.5. Cette figure représente la dispersion de l'estimateur selon le tirage du jeu de données. Pour l'exemple de données présenté (Figure 3.21a), la moyenne des μ_a estimés est égale à -0.4759 et l'écart-type est égale à 0.0964. Pour le paramètre σ_a estimé, les valeurs de la moyenne et de l'écart-type sont égales à 0.3556 et 0.0326 respectivement.

De même, en complément, la Figure 3.21b présente un échantillon suivant la distribution de la vraie valeur du paramètre pour un seul jeu de données. Sa moyenne correspond donc à $\hat{\theta}$ et l'écart-type représente l'incertitude d'estimation. Pour l'exemple étudié (Figure 3.21b), les valeurs de la moyenne et de l'écart-type du paramètre μ_a sont égales à -0.4345 et 0.0593, et pour σ_a , 0.3395 et 0.0487 respectivement. Pour chacun des points



(a) Paramètres estimés avec tirage des durées de vie groupées

(b) Échantillonnage des paramètres selon la loi asymptotique des estimateurs

FIGURE 3.21 – Échantillons de μ_t et σ_t par deux approches

de la Figure 3.21a, cette incertitude d'estimation peut également être représentée.

Dans les deux cas, les moyennes des paramètres μ_a et σ_a des échantillons sont comparables. L'écart-type du paramètre μ_a pour le premier cas est plus important que pour le deuxième, et inversement pour le paramètre σ_a .

Dans la suite, on s'intéresse à la variabilité de tirage du jeu de données et des données groupées correspondant (Figure 3.21a).

3.7.2 Réseau de fonctions à base radiale

Pour étudier la variabilité des paramètres μ_a, σ_a identifiés par les méthodologies proposées, une analyse de régression est réalisée à l'aide d'un réseau de fonctions à base radiale [97]. Le réseau de fonctions à base radiale (RBF) est un réseau de neurones artificiels. Il est constitué en général de deux couches, une première couche de fonction à base radiale et une deuxième couche linéaire présentée dans la Figure 3.22. Ceci signifie que la sortie du réseau est une combinaison linéaire de fonctions à base radiale dépendant des entrées et des paramètres des neurones. Dans la Figure 3.22, il y a N données d'entrée \underline{x} et N données de sortie \underline{y} , notées $\mathbf{x} = \underline{x}_1, \dots, \underline{x}_N$ et $\mathbf{y} = \underline{y}_1, \dots, \underline{y}_N$. M couches cachées sont basées sur RBF de centre $\mathbf{c} = \underline{c}_1, \dots, \underline{c}_M$. La sortie est une combinaison linéaire de RBF avec des paramètres $\mathbf{w} = w_1, \dots, w_M$.

La fonction à base radiale est une fonction dont la valeur dépend seulement de la distance entre le point de calcul et le centre, cette distance peut être représentée par la distance euclidienne. Par exemple, deux points e et e' de coordonnées respectives (x, y) et (x', y') , leur distance euclidienne est $d = \|e - e'\| = \sqrt{(x - x')^2 + (y - y')^2}$. Différentes fonctions à base radiale Φ existent, par exemple la fonction Gaussienne $\Phi(d) = \exp(-(\gamma d)^2)$ ou la fonction multiquadratique $\Phi(d) = 1 + (\gamma d)^2$. La fonction Gaussienne est souvent utilisée dans ce type de réseau. M centres sont demandés dans M couches cachées constituées de RBF, et aussi les largeurs γ de la fonction Gaussienne. La

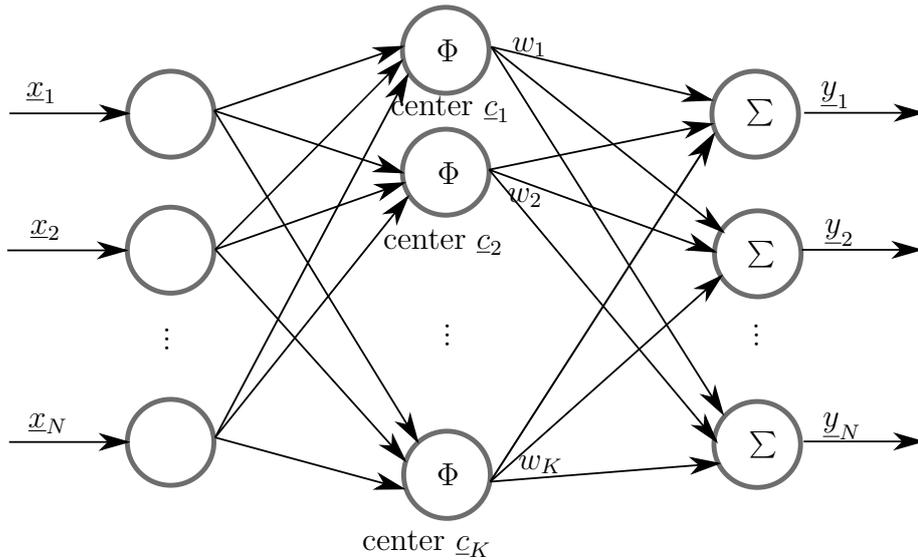


FIGURE 3.22 – Schéma du réseau de fonctions à base radiale

la \$j^{i\text{me}}\$ couche cachée de RBF est présentée par

$$\Phi(\underline{x}) = \exp(-\gamma_i \|\underline{x} - \underline{c}_i\|^2), i \in \llbracket 1, M \rrbracket. \quad (3.18)$$

Chaque fonction a un support à symétrie sphérique centré en \$\underline{c}_i\$ et de taille caractérisée par \$\gamma_i\$. Une donnée contribuera à la valeur de \$w_i\$ si elle est dans le voisinage de \$\underline{c}_i\$, ce qui confirme un caractère local de la surface de réponse.

La forme standard pour calculer la sortie avec un réseau RBF s'écrit,

$$h(\underline{x}) = \sum_{i=1}^M w_i \exp(-\gamma_i \|\underline{x} - \underline{c}_i\|^2), i \in \llbracket 1, M \rrbracket, \quad (3.19)$$

où les paramètres à déterminer sont les centres et largeur de la fonction Gaussienne ainsi que le poids \$w_1, \dots, w_p\$ pour la combinaison des couches cachées RBF.

Deux types de réseau RBF et la détermination de paramètres sont maintenant présentés.

3.7.2.1 Réseau RBF exact

Dans un réseau RBF exact, il y a autant de couches cachées que de données d'entrée, c'est-à-dire \$M = N\$. Les centres des RBF sont choisis pour des positions correspondant aux données d'entrée, le réseau RBF s'écrit

$$h(\underline{x}) = \sum_{i=1}^N w_i \exp(-\gamma_i \|\underline{x} - \underline{c}_i\|^2), i \in \llbracket 1, M \rrbracket, \quad (3.20)$$

Dans ce cas, les données sont divisées en \$N\$ grappes, l'erreur aux points de données est nulle.

Les paramètres \$w_1, \dots, w_M\$ sont déterminés par la résolution d'un système d'équations linéaires. En notant \$\Phi_{ij} = \exp(-\gamma_i \|\underline{x}_j - \underline{c}_i\|^2), i \in \llbracket 1, M \rrbracket, j \in \llbracket 1, N \rrbracket\$ avec \$M = N\$, on a

$$\begin{bmatrix} \Phi_{11} & \Phi_{21} & \cdots & \Phi_{N1} \\ \Phi_{12} & \Phi_{22} & \cdots & \Phi_{N2} \\ & \vdots & & \\ \Phi_{1N} & \Phi_{2N} & \cdots & \Phi_{NN} \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_N \end{bmatrix} = \begin{bmatrix} \underline{y}_1 \\ \underline{y}_2 \\ \vdots \\ \underline{y}_N \end{bmatrix}, \quad (3.21)$$

alors $\mathbf{w} = \Phi^{-1}\mathbf{y}$.

Un réseau RBF exact permet de créer un réseau interpolant, c'est-à-dire que les sorties du réseau sont égales aux données, aux points de données. Lorsque la quantité de données d'entrée et de sortie est importante, le nombre de couches cachées dans le réseau devient trop grand, une telle richesse peut ne pas être utile. Il est alors possible de représenter les N données par M , $M < N$, centres de grappe, le réseau s'appelle un réseau RBF approximant.

3.7.2.2 Réseau RBF approximant

Dans le réseau RBF approximant, M couches cachées sont utilisées pour N données d'entrée avec $M < N$, la forme du réseau devient

$$h(\underline{x}) = \sum_{i=1}^M w_i \exp(-\gamma_i \|\underline{x} - \underline{c}_i\|^2). \quad (3.22)$$

L'apprentissage d'un réseau RBF consiste à déterminer le nombre de couches cachées M et les paramètres, les centres c des RBF, la largeur γ des fonctions Gaussiennes et les poids w des RBF. La fonction objectif d'apprentissage est l'erreur de modèle (erreur quadratique) aux points de données

$$E = \frac{1}{2} \sum_{j=1}^N (h(\underline{x}_j) - \underline{y}_j)^2, j \in \llbracket 1, N \rrbracket. \quad (3.23)$$

Pour déterminer les paramètres, on génère souvent les valeurs initiales aléatoirement, puis l'apprentissage des valeurs est réalisé. Différentes méthodes d'apprentissage existent, par exemple la méthode de Back Propagation supervisée, la méthode hybride qui combine une partie non supervisée et une supervisée proposée par [98]. Cette méthode est intéressante en termes de simplicité, de précision et de temps de calcul. Une méthode hybride est ensuite présentée.

3.7.2.2.1 Apprentissage des centres c du RBF

Cet apprentissage peut être réalisé par une méthode non supervisée, c'est-à-dire sans utiliser les données de sortie. Ici la méthode des K -moyennes est présentée.

L'objectif est de minimiser la distance entre chaque donnée d'entrée \underline{x} et le centre le plus proche. Supposons que les données d'entrée soient divisées en M grappes S_1, \dots, S_M de centres c_1, \dots, c_M , on veut minimiser

$$\sum_{i=1}^M \sum_{\underline{x}_j \in S_i} \|\underline{x}_j - \underline{c}_i\|^2, j \in \llbracket 1, N \rrbracket.$$

Il y a deux paramètres à optimiser, les positions des centres \underline{c}_i et le choix des données d'entrée de chaque grappe S_i . Ce problème peut être traité par l'algorithme de Lloyd.

L'idée principale de cet algorithme est de faire une minimisation itérative, tout d'abord en fixant les grappes, les centres sont calculés par la moyenne de chaque grappe

$$\underline{c}_i = \frac{1}{|S_i|} \sum_{\underline{x}_j \in S_i} \underline{x}_j, \forall i \in \llbracket 1, M \rrbracket$$

ensuite avec les \underline{c}_i obtenus, S_i est mis à jour en prenant tous les points dont \underline{c}_i est le plus proche centre

$$S_i = \{\underline{x}_j : \|\underline{x}_j - \underline{c}_i\| \leq \|\underline{x}_j - \underline{c}_l\|, \forall l \in \llbracket 1, M \rrbracket, l \neq i\}.$$

Pour effectuer cet algorithme, il faut une valeur initiale pour les \underline{c}_i et S_i , le résultat converge mais possiblement vers un minimum local. Le résultat dépend fortement des valeurs initiales, il est donc conseillé d'essayer plusieurs valeurs initiales et d'évaluer les minimums finaux.

Cette façon d'apprentissage dépend seulement des données d'entrée, donc c'est une méthode non supervisée.

3.7.2.2.2 Apprentissage de largeur γ du RBF

La largeur γ peut être considérée comme une constante pour toutes les fonctions Gaussiennes [99], ce qui est le cas dans notre travail. Par défaut, la largeur est égale à 1, cela peut être défini par l'utilisateur.

D'autres méthodes ont également été proposées. Moody et Darken [98] ont proposé l'apprentissage de largeur γ des fonctions Gaussiennes par la méthode heuristique de 'P voisins-plus proches'. Le principe est de varier les largeurs afin d'avoir un certain nombre de données recouvrant deux grappes voisines pour qu'il ait une interpolation continue sur les régions d'espace d'entrée. Les heuristiques peuvent être reformulées et minimisées par rapport aux largeurs γ . Benoudjit et al. [100] ont proposé une méthode qui considère un facteur d'échelle de largeur et ont optimisé ce facteur par minimisation de l'erreur quadratique moyenne du réseau. Lazaro et al. [101] ont utilisé l'algorithme espérance-maximisation pour accélérer l'apprentissage des paramètres dans une façon supervisée.

3.7.2.2.3 Apprentissage des pois w

Une fois les couches cachées fixées, la sortie du réseau se calcule comme une combinaison linéaire des M couches cachées du RBF, l'Équation (3.21) devient

$$\begin{bmatrix} \Phi_{11} & \Phi_{21} & \cdots & \Phi_{M1} \\ \Phi_{12} & \Phi_{22} & \cdots & \Phi_{M2} \\ & \vdots & & \\ \Phi_{1N} & \Phi_{2N} & \cdots & \Phi_{MN} \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_M \end{bmatrix} = \begin{bmatrix} \underline{y}_1 \\ \underline{y}_2 \\ \vdots \\ \underline{y}_N \end{bmatrix}, \quad (3.24)$$

avec $M < N$. Pour résoudre ce système pour lequel il y a plus d'équations que d'inconnues, un estimateur de moindres carrés est souvent utilisé

$$\mathbf{w} = (\Phi^T \Phi)^{-1} \Phi^T \mathbf{y}. \quad (3.25)$$

La détermination du paramètre w est supervisée en utilisant la sortie y . Grâce à la relation linéaire de cette couche, il est facile à déterminer. Une fois w déterminé, la sortie du réseau peut être estimée en n'importe quel point.

3.7.3 Résultat de l'étude sur la précision des méthodologies d'identification

Dans cette étude, les courbes de Wöhler probabilistes sont générées avec les paramètres de référence de Renault pour pouvoir simuler une situation réaliste, et à partir des séquences des amplitudes des contraintes S et nombre de cycles correspondant n compté par la technique de Rainflow, l'échantillon des durées de vie est généré. Pour simuler les données expérimentales groupées, l'échantillon est distribué dans les intervalles de temps dont les bornes sont $\{0; 50\%; 100\%; 150\%; 200\%; \infty\}$. 100 échantillons de durée de vie chacun de taille 40 sont obtenus, la loi de distribution Log-normale est estimée pour ces échantillons. 100 paires des paramètres (μ_t, σ_t) sont pris comme l'entrée des méthodologies d'identification, avec les fonctions de coût CF1, CF3 et CF4, on obtient 100 paires des paramètres (μ_a, σ_a) .

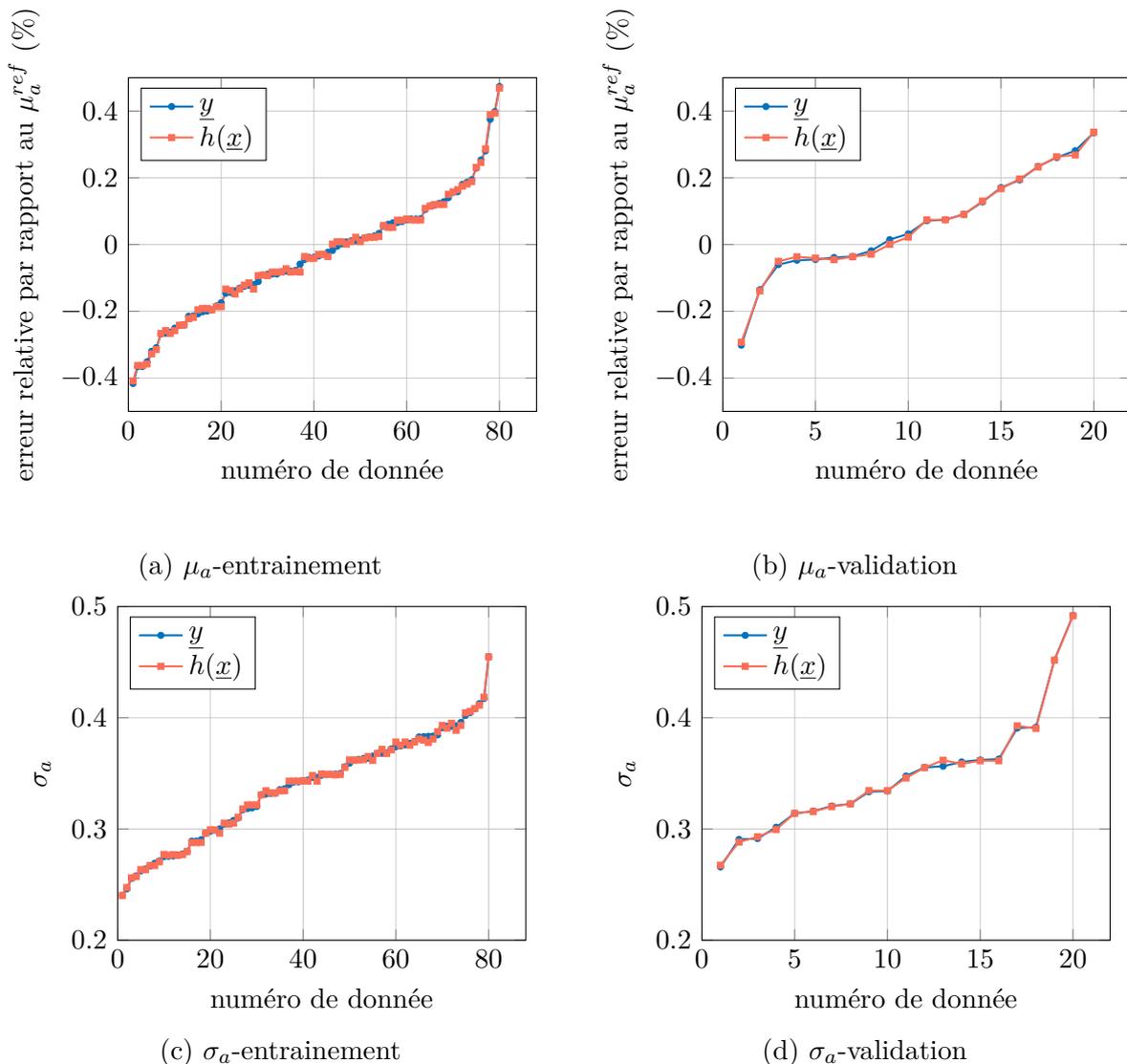


FIGURE 3.23 – Entraînement du RBF pour (a) μ_a et (c) σ_a . Validation du RBF pour (b) μ_a et (d) σ_a avec la fonction de coût CF1

Avec 80 paires de (μ_t, σ_t) et (μ_a, σ_a) , un réseau RBF est entraîné, ce réseau est

vérifié avec les 20 autres paires de données. Pour les fonctions de coût CF1, CF3 et CF4, l'entraînement et la validation du réseau sont présentés dans les Figures 3.23, 3.25 et 3.26. Pour des raisons de confidentialité, la valeur de référence du paramètre μ_a ne sera pas présentée, les paramètres μ_a identifiés seront présentés en erreur relative par rapport à la valeur de référence.

Pour CF1, la comparaison entre les données de μ et σ par minimisation de la fonction de coût et la sorties du RBF sont présentées sur la figure 3.23. Les 80 données pour entraîner le RBF sont présentées à gauche de la Figure 3.23(a et c), les 20 données pour vérifier RBF sont présentées à droite Figure 3.23(b et d). Les paramètres prédits par le réseau RBF et les valeurs de sortie sont assez proches.

L'évolution de l'entraînement du RBF pour CF1 est présentée (Figure 3.24), il s'agit de la somme de l'erreur quadratique entre les données identifiées par les méthodologies et la sortie prévue par RBF en fonction du nombre de couches cachées. En fait, l'erreur diminue rapidement pour les trois premières couches de $2.943e^{-3}$ à $5.539e^{-6}$, puis diminue lentement jusqu'à 80 couches correspondant au nombre de données utilisées pour l'entraînement du réseau. Finalement, pour 80 couches cachées, l'erreur est égale à $4.967e^{-6}$. Vu l'évolution de l'erreur, 4 couches cachées semblent suffisantes pour ce RBF. Il est également remarqué que le réseau 'exact' avec 80 couches n'a pas donné une interpolation exacte, car il y a répétition des valeurs des données (voir section 2.2) du fait du groupement des durées de vie. En revanche, les sorties (paramètres identifiés avec les fonctions de coût) des mêmes valeurs de données d'entrée sont différentes. La matrice dans l'Équation (3.21) n'est plus de rang complet ($M > N$), ce système ne peut normalement pas être résolu. Le réseau est ici obtenu par optimisation de l'erreur du réseau. Selon les résultats pour tester le réseau, il est montré que le réseau peut donner une prédiction plutôt cohérente.

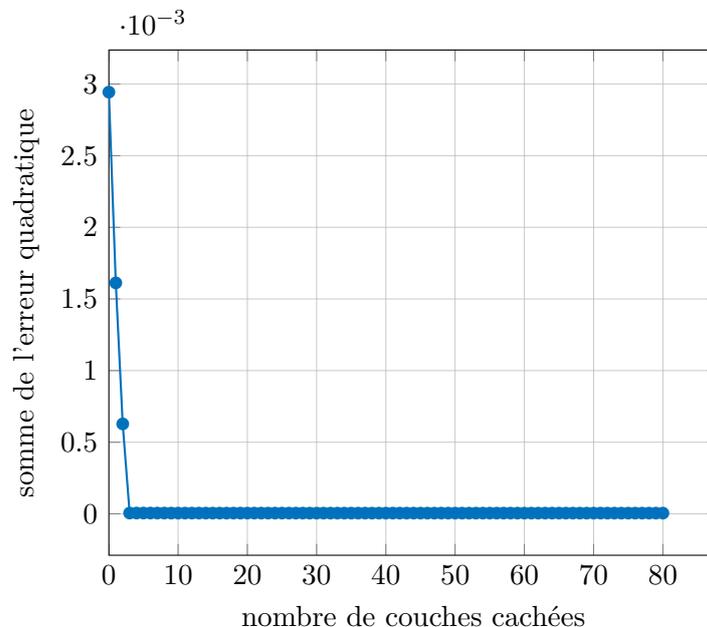


FIGURE 3.24 – Entraînement du RBF pour la fonction de coût CF1

Pour CF3, les données de μ et σ sont présentées dans leur ordre croissant respectif pour mieux observer la performance du réseau en fonction de la grandeur des valeurs

dans la Figure 3.25. La valeur des données pour tester le RBF est bien couverte par les données utilisées pour entraîner le RBF. Les données pour tester sont principalement bien prédites, sauf deux points pour chacun des paramètres μ_a et σ_a . La performance du réseau RBF est satisfaisante.

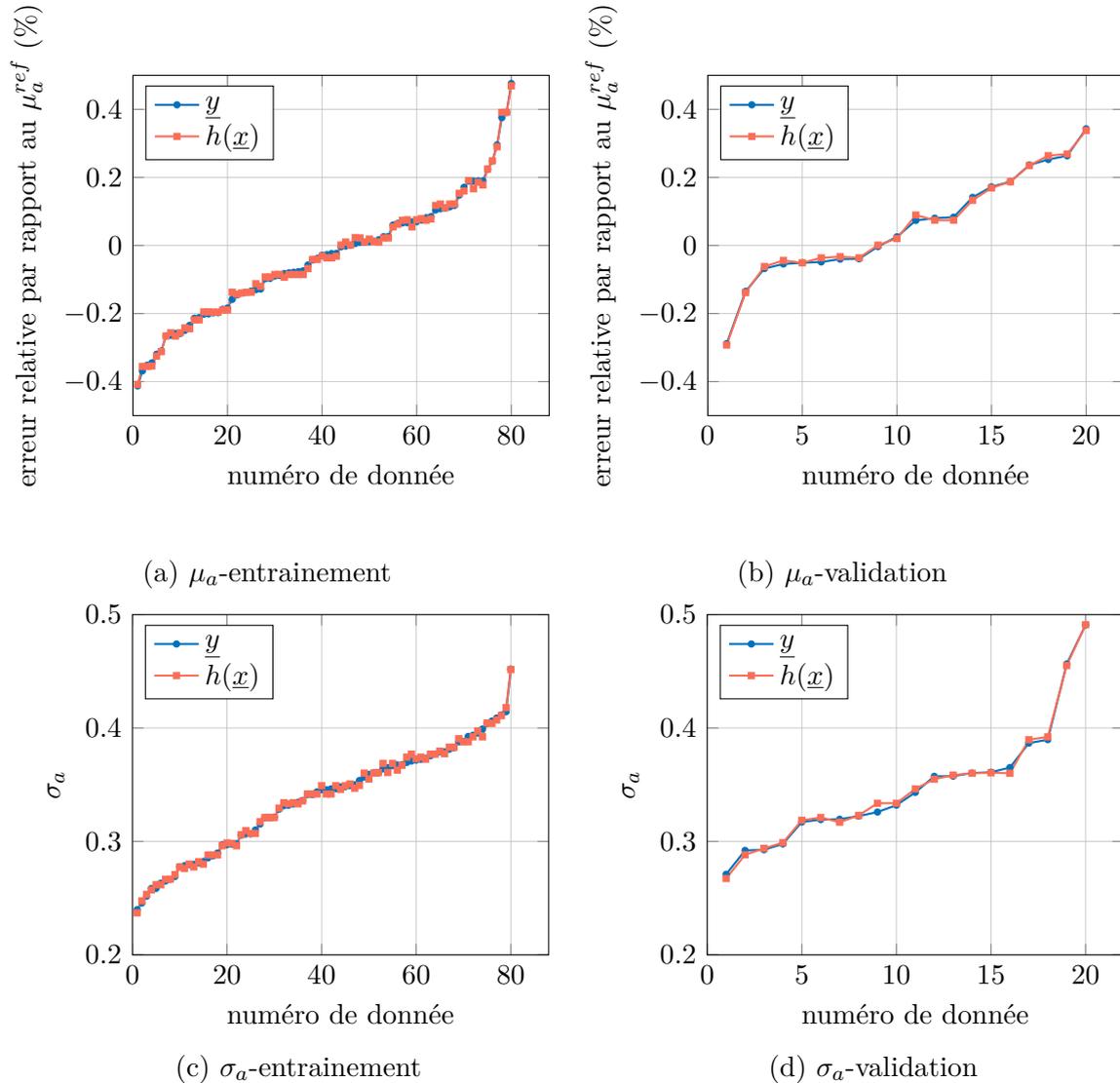


FIGURE 3.25 – Entraînement du RBF pour (a) μ_a et (c) σ_a . Validation du RBF pour (b) μ_a et (d) σ_a avec la fonction de coût CF3

Pour la fonction de coût CF4, les données de μ et σ sont présentées par ordre croissant dans la Figure 3.26. La prédiction du RBF est satisfaisante.

Pour évaluer la performance des réseaux entraînés avec les résultats des différentes fonctions de coût, l'erreur relative entre les sorties du RBF et les vraies valeurs sont calculées, l'erreur relative maximale est présentée dans le Tableau 3.7. Pour les réseaux RBF des 3 fonctions de coût, le paramètre μ_a est globalement mieux prédit que le paramètre σ_a . La performance des RBF est satisfaisante, ils peuvent être utilisés à la place des méthodologies d'identification pour accélérer le processus d'identification.

Pour les 100 paires des paramètres (μ_t, σ_t) estimés à partir des données groupées

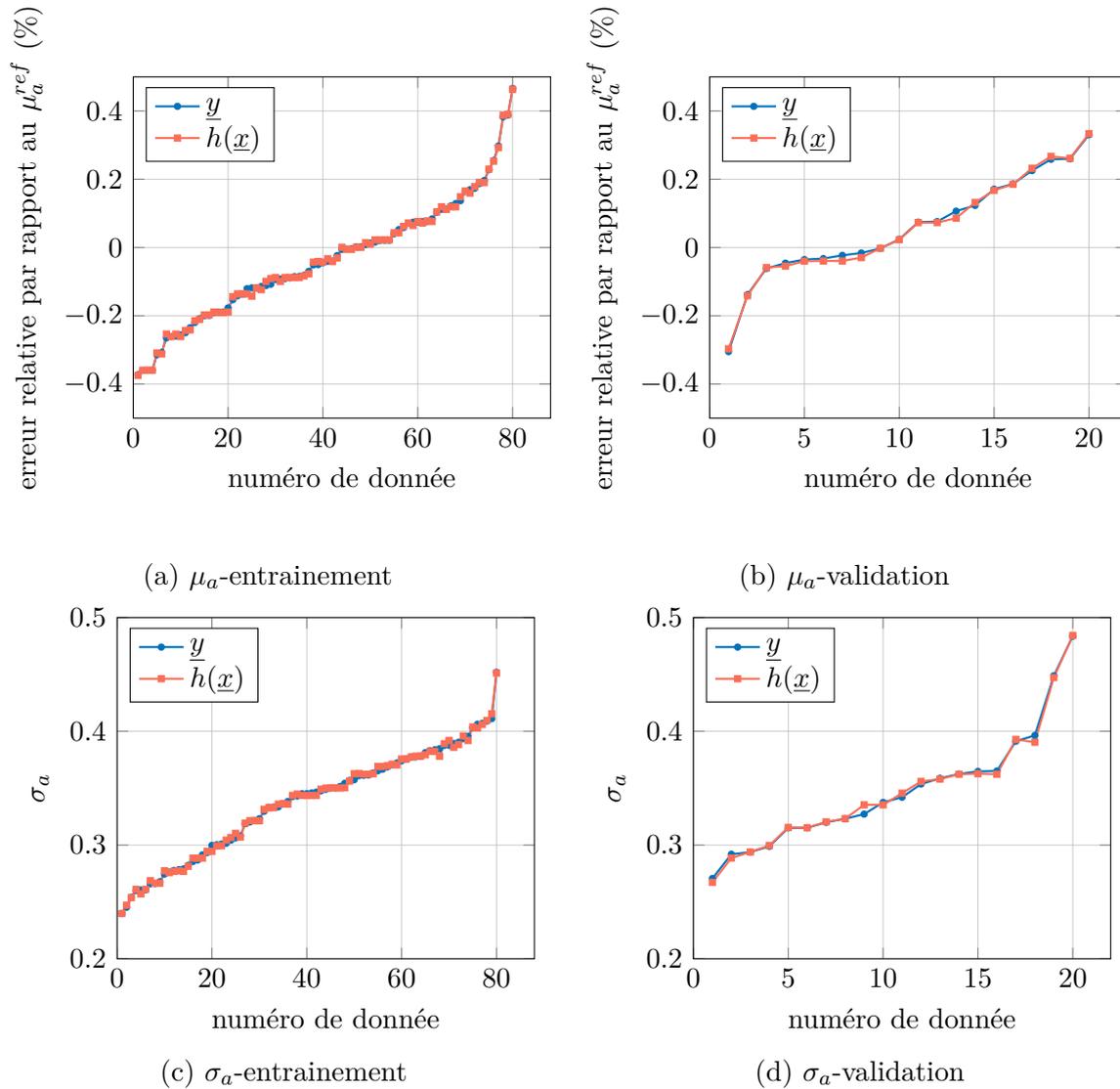


FIGURE 3.26 – Entrainement du RBF pour (a) μ_a et (c) σ_a . Validation du RBF pour (b) μ_a et (d) σ_a avec la fonction de coût CF4

TABLEAU 3.7 – Erreur relative des réseaux RBF pour 3 types de fonction de coût

		CF1	CF3	CF4
Erreur relative	μ_a	0.04%	0.12%	0.05%
	σ_a	3.85%	7.40%	3.42%

générees d'une même loi, les résultats d'identification des paramètres μ_a, σ_a (via la fonction de coût CF1) sont présentés dans la Figure 3.27. Les valeurs du paramètre μ_a sont incluses dans l'intervalle $[-0.421\%, 0.481\%]$ autour de la référence μ_a^{ref} dont la moyenne a une erreur relative égale à -0.0039% par rapport à la valeur référence. L'écart-type est égale à 0.0589. Pour σ_a , les valeurs prédites par RBF sont incluses dans l'intervalle $[-34.286\%, 40\%]$ autour de la valeur de référence $\sigma_a^{ref} = 0.35$, avec une moyenne égale à 0.3408 et un écart-type 0.0495. Ceci permet d'estimer la précision d'identification attendue sur un jeu de données aussi représentatif que possible.

Pour conclure, à partir d'une loi de distribution de la durée de vie, la variabilité du résultat des méthodologies d'identification a été analysée en tenant compte de la variabilité des données à caractère groupé, de l'incertitude de la méthode d'estimation et de la variabilité provenant du tirage. Un réseau à fonctions de base radiale a été construit pour simuler les méthodologies d'identification pour 3 fonctions de coût CF1, CF3 et CF4, la performance est satisfaisante et permet d'accélérer l'évaluation de la méthodologie.

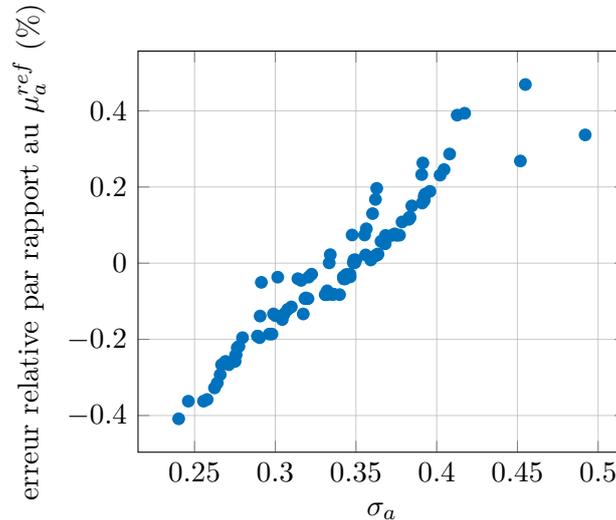


FIGURE 3.27 – Paramètres μ_a et σ_a prédits par le réseau RBF de la fonction coût CF1

3.8 Conclusion

La dispersion de la courbe S-N entraîne une variabilité importante pour le calcul de la durée de vie en fatigue des structures mécaniques. De manière conventionnelle, la dispersion est déterminée par des essais sur éprouvettes. Cependant, cette méthode n'est pas adaptée aux joints de soudure dont les propriétés sont liées au procédé de fabrication de structures potentiellement complexes. La propriété de la courbe S-N des matériaux dans de telles structures peut être très différente de celle obtenue par de simples essais sur éprouvettes. Cette étude a proposé une stratégie d'identification pour un modèle probabiliste de la courbe S-N, à partir de données d'essais de fatigue expérimentaux réels de structures et de données de simulation numérique. La courbe S-N stochastique calibrée peut être utilisée pour les simulations numériques et pour la prédiction de la variabilité de la durée de vie des structures complexes.

Au cours de cette étude, nous avons proposé de nouvelles fonctions de coût pour les problèmes inverses stochastiques afin de mesurer la différence entre deux populations de variables aléatoires, qui sont très générales. Elles peuvent également être utilisées pour d'autres problèmes d'identification. La performance des méthodes proposées a été étudiée avec des exemples numériques et il s'avère que la stratégie globale est robuste et efficace. Les fonctions de coût construites à l'aide de la comparaison paramétrique, du test de Cramér-von Mises à un et à deux échantillons ont de meilleures performances pour la précision de l'identification. Les nouvelles fonctions objectifs proposées fournissent des

outils pour mesurer la différence entre une distribution et un échantillon, ou entre des échantillons. Elles peuvent être utilisées dans l'identification stochastique pour d'autres problèmes inverses.

Chapitre 4

Application industrielle

Dans ce chapitre, l'application des travaux de thèse au contexte industriel est présentée.

Contents

4.1 Introduction	116
4.2 Identification des paramètres probabilistes de la courbe de Wöhler pour deux essieux	116
4.2.1 Présentation des essieux X52 et X98	116
4.2.1.1 Essieu X52	116
4.2.1.2 Essieu X98	117
4.2.2 Résultats d'identification des paramètres probabilistes de la courbe de Wöhler	118
4.2.2.1 Résultats pour l'essieu X52	119
4.2.2.2 Résultats pour l'essieu X98	120
4.2.2.3 Comparaison des résultats de X52 et X98	121
4.3 Exploitation des résultats de la thèse	122
4.3.1 Méthode d'estimation	122
4.3.2 Propagation des incertitudes sur la durée de vie	122
4.3.3 Critère de corrélation	124
4.4 Conclusions et perspectives	125

4.1 Introduction

Ce projet de recherche a notamment été initié car il était constaté que la simulation numérique de l'essai synthétique menait à une durée de vie déterministe, issue du calcul par éléments finis, qui n' était pas représentative de la durée de vie expérimentale dispersée, observée lors de l'essai physique. Des méthodologies d'identification ont donc été développées dans la thèse pour permettre de construire un dialogue entre essai et calcul (Figure 1.17) pour améliorer la performance de la simulation numérique. Des méthodes d'estimation des durées de vie expérimentales ont également été proposées pour alimenter ce dialogue. Cette approche repose sur deux hypothèses. La première suppose que la dispersion observée lors des essais, soit la variabilité de la durée de vie, est principalement induite par les incertitudes liées à la courbe de Wöhler, autrement dit qu'elle peut se formaliser dans la courbe de Wöhler. La seconde hypothèse considère que le modèle numérique développé par Renault de châssis est représentatif. Le schéma du dialogue entre essai et calcul inclut les méthodes d'estimation des durées de vie expérimentales présentées dans le chapitre 2 et les méthodologies d'identification analytique et par simulation introduites dans le chapitre 3.

Dans ce chapitre, les méthodologies d'identification sont tout d'abord appliquées sur les essieux Renault X52 et X98. Cette démarche vise à identifier la courbe de Wöhler pour chaque cordon de soudure, qui devrait être similaire pour chaque essieu de même type. Plus précisément, l'objectif est de quantifier la variabilité liée à la courbe de Wöhler pour mieux interpréter les résultats des essais synthétiques sur le châssis en terme de durée de vie. Puis, les méthodes développées et les résultats obtenus dans cette thèse sont exploités afin d'être appliqués sur des cas industriels. En effet, la méthode d'estimation des données groupées contribue au traitement des données expérimentales issues des essais. La courbe de Wöhler probabiliste identifiée aide à estimer la durée de vie probabiliste et à proposer un intervalle de confiance pour la confrontation avec les résultats expérimentaux. Dans ce but, une interface graphique a été développée sous MATLAB (Annexe F) afin d'évaluer la variabilité sur les projets Renault.

4.2 Identification des paramètres probabilistes de la courbe de Wöhler pour deux essieux

Les essieux X52 Ph1 Ph2 et X98 sont présentés, puis les résultats d'identification des paramètres probabilistes de la courbe de Wöhler sont proposés et discutés.

4.2.1 Présentation des essieux X52 et X98

4.2.1.1 Essieu X52

L'essieu arrière X52 est un essieu pour le véhicule Sandero de Dacia. La Figure 4.1 le présente avec le système de suspension. Deux itérations de conception, les phases 1 (Ph1) et 2 (Ph2), sont ici considérées. Pour ces deux phases, les géométries de la structure sont identiques, mais le cahier des charges (CDC) est différent (section 1.1.1). L'essieu est



FIGURE 4.1 – Essieu arrière X52

composé de 26 paires de cordons de soudure de chaque côté de la structure symétrique. Les essais expérimentaux synthétiques et les simulations numériques mettent en évidence que 10 paires de cordons mènent à la défaillance. L'emplacement de ces cordons est représenté sur la Figure 4.2, les couleurs bleue et rose indiquent respectivement les cordons situés au premier plan et en arrière plan de la figure. Les méthodologies d'identification des paramètres probabilistes de la courbe de Wöhler sont appliquées sur ces 10 paires de cordon de soudure pour l'essieu X52 Ph 1 et l'essieu X52 Ph 2.

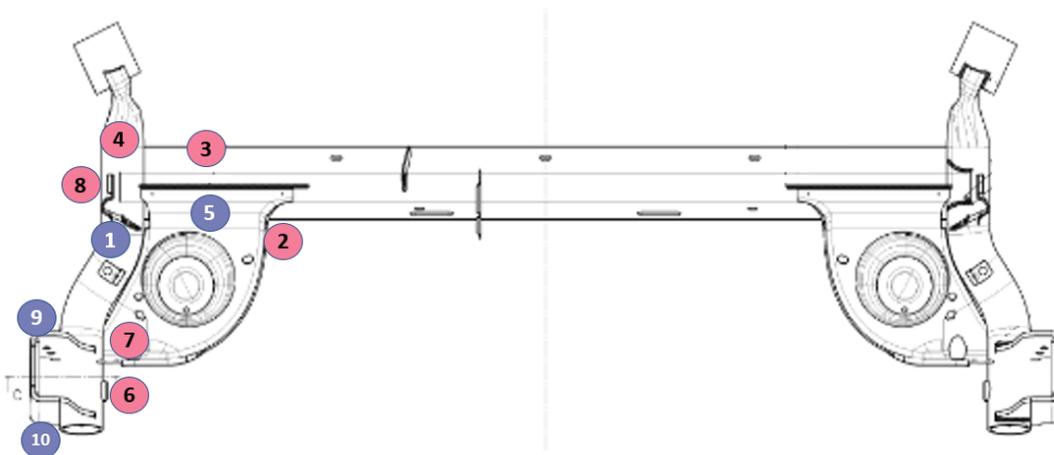


FIGURE 4.2 – Essieu X52 et emplacements des paires de cordons de soudure étudiés

4.2.1.2 Essieu X98

L'essieu X98 est un essieu souple arrière de véhicule Clio 4 présenté dans la Figure 4.3. Cet essieu est composé de 37 paires de cordons de soudure, parmi lesquelles 6 paires



FIGURE 4.3 – Essieu arrière X98

sont considérées défaillantes à l'issue des essais synthétiques et des simulations. L'emplacement de ces cordons est représenté sur la Figure 4.4, les couleurs bleue et rose indiquent respectivement les cordons situés au premier plan et en arrière plan de la figure. Les méthodologies d'identification des paramètres probabilistes de la courbe de Wöhler sont appliquées sur ces 6 paires de cordon de soudure pour l'essieu X98.

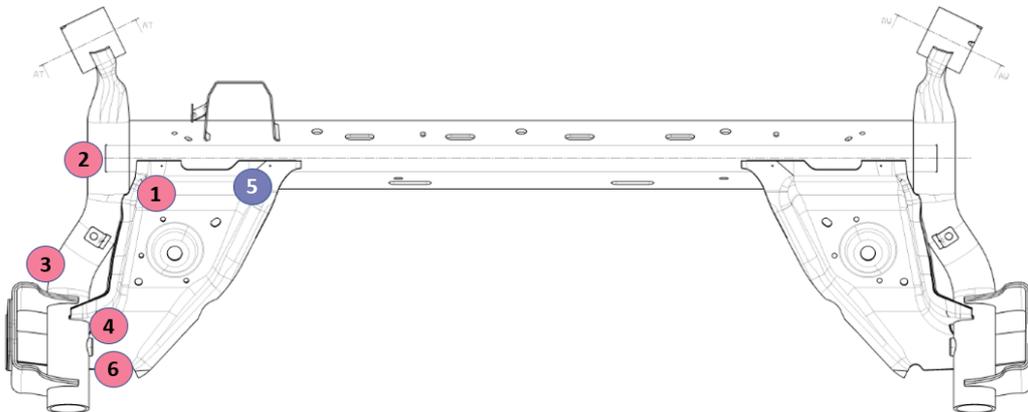


FIGURE 4.4 – Essieu X98 et emplacement des paires de cordons de soudure étudiés

4.2.2 Résultats d'identification des paramètres probabilistes de la courbe de Wöhler

Dans la conception des véhicules Renault, une valeur de référence est utilisée pour le paramètre μ_a de la courbe Wöhler (section 1.2.3). Le paramètre μ_a représente la coordonnée à l'origine de la courbe de Wöhler sur l'axe $\log(N)$ dans le système de coordonnées logarithmique. Le paramètre σ_a représente le niveau de dispersion du logarithme de la durée de vie $\log(N)$, il n'existe pas de valeur de référence pour ce paramètre. Pour des raisons de confidentialité, les résultats obtenus par les méthodologies d'identification sont présentés sous la forme de la valeur absolue d'un écart relatif e_{μ_a} par rapport à la valeur de référence pour le paramètre μ_a , .

Lors des simulations numériques menant à l'évaluation de l'endommagement sur l'essieu X52 Ph1, les résultats sont parfois non symétriques pour une géométrie de structure symétrique et un maillage symétriques. Cela peut être causé par l'asymétrie dans le cahier des charges. De ce fait, les données expérimentales et celles de la simulation numérique sont traitées différemment pour des cordons dits symétriques ou non-symétriques. Les données expérimentales des deux extrémités de la structure sont ainsi prises en compte comme une unique zone pour les cordons symétriques, mais traitées séparément pour les cordons non-symétriques. Pour les résultats de l'identification suivants, les zones présentant un cordon de soudure sont numérotées : les lettres A et B sont ajoutées lorsque il n'y a pas de symétrie, par exemple 1A pour le cordon 1.

4.2.2.1 Résultats pour l'essieu X52

Les zones de cordon de soudure pour lesquelles aucune défaillance n'est observée ne sont pas traitées (par exemple, les zones 1, 6 et 10 pour l'essieu X52 Ph1). Pour le projet X52 Ph1 et Ph2, les résultats de l'identification des paramètres probabilistes de la courbe de Wöhler sont respectivement présentés dans le Tableau 4.1 et le Tableau 4.2. La méthodologie d'identification analytique est notée 'Analytique'. La méthodologie d'identification par l'approche simulation avec la fonction coût paramétrique (CF1) est notée 'Paramétrique', la statistique du test de Kolmogorov-Smirnov d'un échantillon et la statistique du test de Cramér-von Mises d'un échantillon sont notées 'test de KS' (CF3) et 'test de CVM'(CF4) respectivement. Les méthodes correspondantes ont été introduites dans le Chapitre 3 (Tableau 3.1).

TABLEAU 4.1 – Résultats de l'identification des paramètres probabilistes de la courbe de Wöhler pour l'essieu X52 Ph1

Zone	Méthodes		Analytique		Paramétrique		Test de KS		Test de CVM	
	e_{μ_a}	σ_a	e_{μ_a}	σ_a	e_{μ_a}	σ_a	e_{μ_a}	σ_a	e_{μ_a}	σ_a
2	2.52%	0.321	2.50%	0.318	2.50%	0.321	2.50%	0.316		
3	4.76%	0.275	4.75%	0.275	4.78%	0.278	4.74%	0.278		
4	3.60%	0.301	3.59%	0.296	3.62%	0.302	3.58%	0.299		
5	8.09%	0.019	8.09%	0.019	8.44%	0.019	8.09%	0.019		
7A	4.94%	0.171	4.94%	0.173	4.93%	0.172	4.94%	0.168		
8A	3.76%	0.835	3.72%	0.823	3.69%	0.826	3.73%	0.859		
8B	5.14%	0.600	5.15%	0.596	5.15%	0.599	5.13%	0.608		
9	2.39%	0.426	2.40%	0.420	2.39%	0.426	2.45%	0.449		

Il est remarqué que les résultats d'identification obtenus (écart relatif e_{μ_a} par rapport à la valeur de référence des paramètres μ_a et σ_a) sont globalement comparables pour les différentes méthodologies d'identification utilisées. Pour rappel les approches utilisées sont l'approche analytique et l'approche simulation avec 3 fonctions coût différentes. Cela valide la performance des fonctions coût introduites dans le Chapitre 3. Concernant la comparaison entre la valeur du paramètre μ_a identifiée et celle de référence chez Renault,

4.2. Identification pour deux projets

TABLEAU 4.2 – Résultats de l'identification des paramètres probabilistes de la courbe de Wöhler pour l'essieu X52 Ph2

Zone	Méthodes		Analytique		Paramétrique		Test de KS		Test de CVM	
	e_{μ_a}	σ_a	e_{μ_a}	σ_a	e_{μ_a}	σ_a	e_{μ_a}	σ_a	e_{μ_a}	σ_a
1	5.83%	0.246	5.80%	0.250	5.82%	0.250	5.82%	0.248		
2	4.41%	0.351	4.41%	0.349	4.41%	0.355	4.40%	0.354		
3	3.24%	0.382	3.26%	0.380	3.23%	0.382	3.27%	0.392		
5	7.72%	0.314	7.74%	0.316	7.72%	0.314	7.73%	0.314		
6	6.10%	0.427	6.09%	0.423	6.10%	0.425	6.11%	0.334		
7	4.11%	0.326	4.11%	0.330	4.11%	0.320	4.10%	0.326		
8	2.81%	0.397	2.80%	0.401	2.81%	0.404	2.80%	0.397		
10A	2.41%	0.963	2.40%	0.988	2.36%	0.954	2.33%	0.962		

les écarts relatifs sont compris entre 2.39% et 8.44% pour l'essieu X52 Ph1, et compris entre 2.33% et 7.74% pour l'essieu X52 Ph2. Ce résultat montre que les méthodologies proposées parviennent à bien identifier le paramètre μ_a . Le paramètre σ_a représente le niveau de dispersion de la courbe de Wöhler probabiliste. Pour l'essieu X52 Ph1 et Ph2, les valeurs identifiées pour le paramètre σ_a pour les cordons de soudure symétriques, sont inférieures à 0.5, la plupart des valeurs sont comprises entre 0.2 et 0.5. Pour les cordons de soudure considérées non-symétriques, le paramètre σ_a prend des valeurs plus élevées. Ce fait peut être lié au nombre de données, les cordons de soudure symétriques ont en effet 2 fois plus de données que ceux non-symétriques.

4.2.2.2 Résultats pour l'essieu X98

Pour le projet X98, les résultats de l'identification des paramètres probabilistes sont présentés dans le Tableau 4.3.

TABLEAU 4.3 – Résultats de l'identification des paramètres probabilistes de la courbe de Wöhler pour l'essieu X98

Zone	Méthodes		Analytique		Paramétrique		Test de KS		Test de CVM	
	e_{μ_a}	σ_a	e_{μ_a}	σ_a	e_{μ_a}	σ_a	e_{μ_a}	σ_a	e_{μ_a}	σ_a
1	5.27%	1.296	5.32%	1.275	5.42%	1.301	5.24%	1.335		
2	3.11%	1.815	3.15%	1.801	3.09%	1.798	3.20%	1.832		
4	2.87%	0.491	2.86%	0.491	2.87%	0.493	2.90%	0.501		
5	3.80%	0.340	3.79%	0.341	3.80%	0.339	3.81%	0.340		
6	3.40%	1.718	3.44%	1.712	3.43%	1.729	3.11%	1.712		

Comme pour l'essieu X52, les résultats d'identification obtenus (écart relatif e_{μ_a} par rapport à la valeur de référence du paramètre μ_a et paramètre σ_a) sont globalement comparables pour les différentes méthodologies d'identification, soient l'approche analytique

et l'approche simulation avec 3 fonctions coût différentes. L'écart relatif entre les valeurs du paramètre μ_a identifiées et celle de référence chez Renault est compris entre 2.87% et 5.27%. Pour le paramètre σ_a qui représente le niveau de dispersion de la courbe de Wöhler, les valeurs sont comprises entre 0.340 et 1.815. Dans les zones d'observations, 6 cordons de soudure sont menés à la défaillance entre 150% et l'infini, et 12 entre 100% et l'infini.

4.2.2.3 Comparaison des résultats de X52 et X98

Pour le projet X52, les résultats de l'identification sont comparables pour les différents cordons de soudure, ce qui n'est pas le cas pour le projet X98. En particulier, les valeurs du paramètre σ_a identifiées pour le projet X52 sont inférieures à 1, la plupart étant inférieures à 0.5. Pour le projet X98, de nombreuses valeurs sont supérieures à 1. En effet, le résultat de l'identification des paramètres probabilistes de la courbe de Wöhler dépend des résultats expérimentaux disponibles en terme d'estimation de durées de vie. Pour le projet X52 Ph1, les résultats de 47 essais sont disponibles, pour les cordons de soudure dits symétriques, soit 94 données. Pour le projet X52 Ph2, 42 résultats d'essais sont disponibles. Pour le projet X98, les résultats de 17 essais sont seulement utilisables, l'estimation statistique des données expérimentales est donc moins précise.

De plus, pour l'essieu X52, 42 essais Ph1 ont un temps de ressuage jusqu'à 150% du temps normalisé et 34 essais Ph2 ont un ressuage jusqu'à 120%. En revanche, pour l'essieu X98, 8 essais sur 17 sont arrêtés avant 100% du temps, le test le plus court se terminant à 58%. La variation des durées des essais, qui peuvent aussi parfois être courtes, a une influence sur l'estimation de la dispersion de la durée de vie, davantage d'information serait nécessaire.

Pour évaluer la variabilité du paramètre a , le coefficient de variation est calculé. La définition du coefficient de variation pour une variable aléatoire est le rapport entre son écart type et sa moyenne. Pour une loi Log-normale $a \sim \mathcal{LN}(\mu_a, \sigma_a)$, le coefficient de variation est défini avec $CoV_a = \sqrt{\exp(\sigma_a^2) - 1}$. Le niveau de variabilité est calculé pour les 2 projets avec les résultats de l'identification par l'approche simulation avec la fonction de coût de comparaison paramétrique (Tableaux 4.4, 4.5 et 4.6). Ces résultats montrent que le niveau de variabilité du paramètre a est très élevé, certaines valeurs aberrantes du coefficient de variation sont même obtenues. Dans ce dernier cas, les données ne permettent pas d'évaluer correctement le niveau de variabilité.

Pour conclure, les paramètres probabilistes identifiés de la courbe de Wöhler sur les deux essieux Renault mènent à des résultats plus fiables pour les projets X52 Ph1 et Ph2 que pour le projet X98. Les valeurs du paramètre σ_a obtenues, comprises entre 0.2 et 0.5, peuvent être considérées comme une référence en terme de dispersion sur le paramètre a de la courbe de Wöhler (Figure 1.16), pour tout essieu similaire à l'essieu X52. Le calcul du coefficient de variation du paramètre a , met en évidence une variabilité significative.

TABLEAU 4.4 – Coefficient de variation du paramètre a pour X52 Ph1

Zone	CoV_a
2	0.327
3	0.281
4	0.302
5	0.019
7A	0.174
8A	0.984
8B	0.653
9	0.439

TABLEAU 4.5 – Coefficient de variation du paramètre a pour X52 Ph2

Zone	CoV_a
1	0.254
2	0.359
3	0.393
5	0.324
6	0.443
7	0.339
8	0.418
10A	1.287

TABLEAU 4.6 – Coefficient de variation du paramètre a pour X98

Zone	CoV_a
1	2.020
2	4.967
4	0.522
5	0.350
6	4.209

4.3 Exploitation des résultats de la thèse

Dans le contexte de la validation des châssis automobiles, les développements réalisés dans cette thèse peuvent être appliqués sous trois aspects par le constructeur automobile Renault :

- l'estimation paramétrique de la densité de probabilité des durées de vie expérimentales,
- la propagation des incertitudes à la durée de vie,
- la corrélation entre les simulations numériques et les essais expérimentaux.

4.3.1 Méthode d'estimation

La méthodologie d'identification des incertitudes nécessite l'estimation des distributions des durées de vie expérimentales de type données groupées. Pour cela, la méthode du maximum de vraisemblance (Chapitre 2) a été utilisée. Pour illustration, le résultat de l'estimation MLE de la durée de vie du cordon de soudure 2 de l'essieu X52 (Figure 4.2) est présenté dans la Figure 4.5. Les données des durées de vie expérimentales groupées sont présentées sur la borne inférieure de chaque intervalle sous la forme d'une barre d'histogramme. La courbe représente la densité de probabilité estimée avec les données groupées. Les nombres de défaillance et de non défaillance sont présentés également sur la figure. Pour comparer l'estimation des données expérimentales et le calcul par simulation, la moyenne de T^{exp} est égale à 261.45%, T^{EF} par la simulation est égale à 235.81%.

4.3.2 Propagation des incertitudes sur la durée de vie

Les travaux proposés pour identifier la variabilité liée à la courbe de Wöhler peuvent être étendus à l'estimation de la variabilité de la durée de vie numérique. Les incertitudes peuvent être en effet propagées via le post-traitement du calcul par éléments finis, qui permet classiquement de calculer le dommage et la durée de vie à partir des contraintes temporelles avec le comptage Rainflow et la courbe de Wöhler. La densité de probabilité de la Figure 4.6 représente cette propagation. Cette densité est comparée avec celle esti-

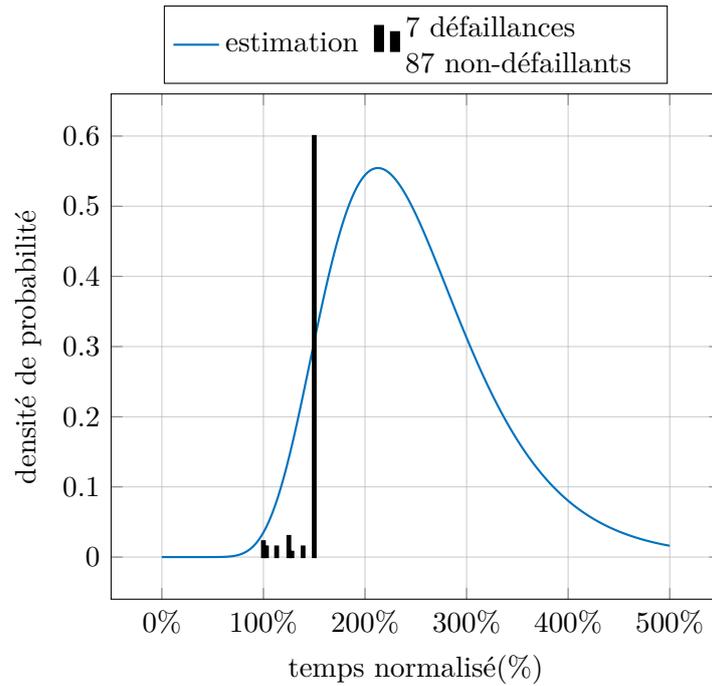


FIGURE 4.5 – Application de la méthode d’estimation du maximum de vraisemblance pour des durées de vie de type données groupées expérimentales

mée des données expérimentales, les courbes sont identiques sur la figure. La densité de probabilité estimée des données expérimentales est donc superposée à celle du calcul probabiliste. La simulation numérique déterministe Renault et la médiane (B50 présenté dans le Chapitre 1) de la distribution de la durée de vie probabiliste obtenue numériquement sont superposées.

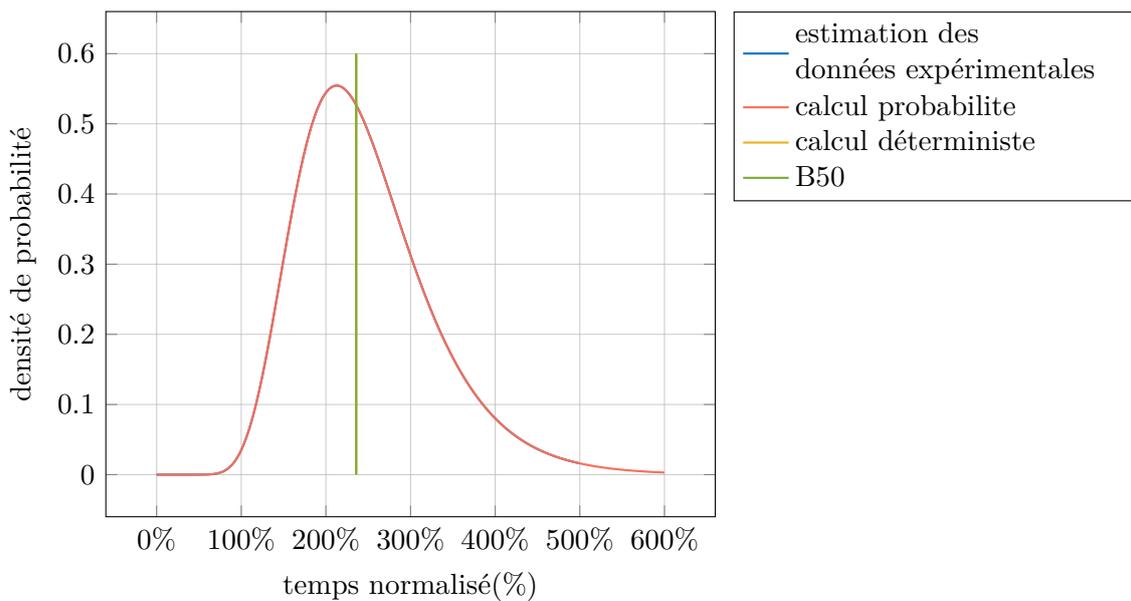


FIGURE 4.6 – Propagation de l’incertitude de la courbe de Wöhler sur la durée de vie

4.3.3 Critère de corrélation

Nous allons présenter une synergie entre les durées de vie expérimentales et celles du calcul numérique pour enrichir la corrélation essai/calcul en terme de critère. La distribution de la durée de vie obtenue par le calcul numérique peut être utilisée pour établir un critère de corrélation entre les simulations numériques et les essais expérimentaux. Suite aux dispersions observées dans les essais, les résultats de la simulation numérique sont considérés acceptables ou pas à l'aide d'un intervalle par rapport aux résultats d'essais. Il est ici proposé un intervalle borné par 5^{ème} percentile de la durée de vie (B5) et 95^{ème} percentile (B95) (Figure 4.7). Si les résultats de la simulation se situent dans cet intervalle, la confiance en ces résultats est de 90%.

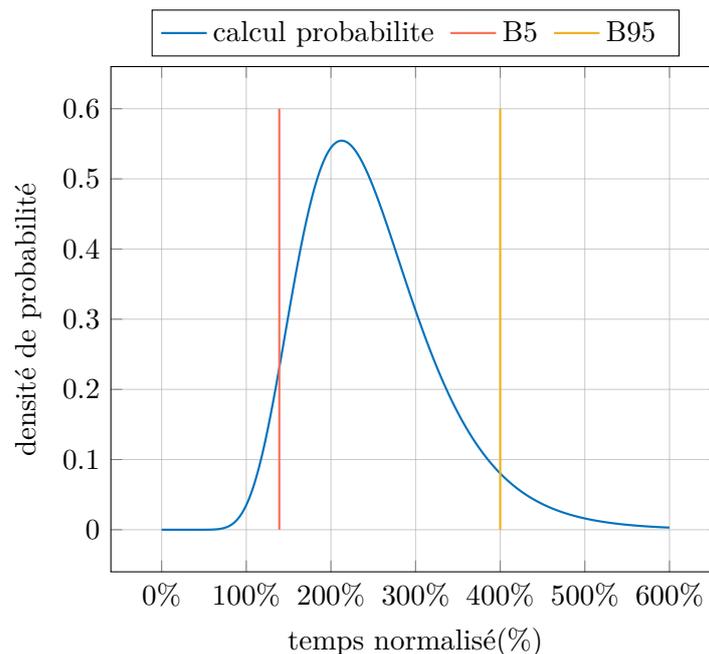


FIGURE 4.7 – Intervalle de confiance à 90% pour le critère de corrélation

Le rapport entre la borne B5 et la médiane B50 est égal à 0.589, le rapport entre la borne B95 et la médiane B50 est égal à 1.696, ces rapports sont interprétés comme des facteurs de corrélation. L'intervalle borné par B5/B50 et B95/B50 donne aussi un critère de corrélation sur le rapport entre le résultat du calcul déterministe et la médiane B50 expérimentale. Ces facteurs de corrélation sont comparables pour différents cordons de soudure du projet X52. Les tableaux 4.7 et 4.8 présentent ces facteurs pour les rapports B5/B50 et B95/B50 pour l'essieu X52 Ph1 et Ph2. En analysant les facteurs obtenus pour le projet X52 Ph1 et Ph2, les facteurs B5/B50 sont compris entre les valeurs 0.2 et 1, les facteurs B95/B50 sont compris entre les valeurs 1 et 4. Pour les cordons de soudure dits symétriques, les valeurs des facteurs B5/B50 sont généralement plus grandes. Alors que pour les cordons de soudure dits non-symétriques, les valeurs des facteurs B95/B50 sont plus petites. La variabilité de la durée de vie des cordons de soudure dits symétrique est donc plus centrée sur la médiane B50. En effet, pour les cordons de soudure symétriques, l'échantillon est plus riche car les deux côtés sont pris en compte en même temps. Les quantités statistiques sont dans ce cas bien estimées. Ces facteurs empiriques peuvent être

utilisés comme un critère de corrélation pour évaluer si le résultat du calcul déterministe est admissible.

TABLEAU 4.7 – Facteurs de corrélation pour l’essieu X52 Ph1

Zone \ Facteur	2	3	4	5	7A	8A	8B	9
B5/B50	0.590	0.636	0.610	0.970	0.754	0.253	0.372	0.496
B95/B50	1.696	1.573	1.640	1.031	1.326	3.952	2.686	2.014

TABLEAU 4.8 – Facteurs de corrélation pour l’essieu X52 Ph2

Zone \ Facteur	1	2	3	5	6	7	8	10A
B5/B50	0.667	0.562	0.532	0.597	0.495	0.585	0.521	0.205
B95/B50	1.498	1.780	1.878	1.676	2.020	1.709	1.921	4.873

4.4 Conclusions et perspectives

Dans ce chapitre, les méthodologies d’identification des paramètres probabilistes de la courbe de Wöhler ont été appliquées sur des essieux de deux projets. L’approche analytique et l’approche par simulation pour différentes fonctions coût ont mené à des résultats comparables. Cela valide les performances des fonctions coût. Les résultats d’identification et les méthodes associées peuvent être exploités par l’industriel en termes de traitement des données expérimentales, de proposition d’un calcul probabiliste de la durée de vie et de proposition d’un critère de corrélation essai-calcul. Ces aspects d’application peuvent aider à analyser les résultats d’essais et également ceux de la simulation numérique. Une interface développée sous MATLAB a été proposée pour faciliter l’utilisation des méthodologies proposées dans la thèse.

Concernant la représentation des données expérimentales, l’industriel utilise la distribution Log-normale, en tenant compte des avis des experts. D’après ce qui a été montré dans la section 2.3.4, la loi Log-normale ne donne pas un taux de défaillance représentatif pour une pièce mécanique. Le taux de défaillance doit généralement être croissant au cours du temps au lieu d’être décroissant comme modélisé par la loi Log-normale. L’utilisation de la loi Log-normale ne serait donc pas suggérée pour modéliser la durée de vie de pièces mécaniques pour estimer le taux de défaillance.

Enfin, des perspectives sont à envisager. L’estimation de la fiabilité est fortement influencée par le nombre d’essais réalisés [10]. Pour les structures prototypes utilisées dans l’essai synthétique, le niveau de dispersion peut être estimé pour l’opération de soudage ou de découpage. La taille de la base de données doit être suffisante, les données de précédents projets et de pièces fabriquées en série peuvent être utilisées. Il serait donc intéressant pour l’industriel de quantifier l’incertitude pour un type de pièce d’une classe d’essieu dans le cadre de l’essai synthétique d’endurance, soit l’incertitude sur la durée de vie dans une zone avec cordons de soudure. Pour cela, il faudra appliquer les méthodologies

d'identification proposées sur davantage de données projets de l'essai synthétique, soit un grand nombre d'essais, afin de pouvoir quantifier les incertitudes dans les zones d'intérêt. Ces informations pourront servir ensuite dans les futurs projets afin de prédire les incertitudes dans les zones d'intérêt d'un type de pièce, avec un nombre d'essais synthétiques raisonnable. Ceci a pour objectif de mieux maîtriser la dispersion des résultats des essais synthétiques en visant à déduire leur nombre. Cette approche d'identification pourrait également être envisagée pour les essais dynamiques, représentatifs de l'essai de roulage sur piste et permettant de valider en endurance les sous-systèmes comme une caisse en blanc.

Conclusion générale et perspectives

Dans cette conclusion générale, le contexte industriel et les développements de la thèse sont rappelés, une synthèse sur les méthodes proposées et les résultats obtenus est présentée. Des perspectives clôturent ce manuscrit.

Dans le contexte industriel du constructeur automobile Renault, la validation des châssis en fatigue combine des essais expérimentaux et des simulations numériques pour vérifier la fiabilité des conceptions. Les essais expérimentaux sont effectués sur différents niveaux de structures suivant un plan d'essais accélérés. Une analyse statistique est nécessaire pour analyser la fiabilité des essais. Les essais expérimentaux d'endurance pratiqués sur les structures automobiles utilisent une méthode d'inspection, le ressuage, permettant de localiser l'amorce de fissures. Les données issues de ce contrôle sont censurées, elles sont de type groupé. Le premier objectif de la thèse était de proposer des méthodes d'estimation adaptées dans le modèle de fiabilité afin de déterminer la distribution des durées de vie expérimentales. Le second objectif des travaux de thèse visait à prédire la variabilité de la durée de vie par la simulation numérique.

Le premier objectif correspond aux développements du chapitre 2, deux méthodes d'estimation et trois modèles de distribution ont été proposés afin d'estimer les durées de vie expérimentales groupées et de quantifier la variabilité. Les méthodes d'estimation sont la méthode du maximum de vraisemblance (MLE) et la méthode Bayésienne. Les modèles de distribution sont les lois de Weibull, Log-normale et multinomiale. La méthode MLE extrait l'information à partir des observations expérimentales, le paramètre estimé est une variable déterministe. Lorsque la quantité des données est suffisante, la méthode MLE est simple et fiable pour estimer la distribution de la durée de vie. L'estimation Bayésienne prend en compte l'information des données expérimentales, mais également une loi *a priori* issue d'avis d'experts ou d'anciennes études. Le paramètre à estimer est une variable aléatoire, les données et la loi *a priori* servent à calculer sa loi *a posteriori*. Le choix de la loi *a priori* est un enjeu important dans l'approche Bayésienne. Les expériences passées peuvent contribuer à ce choix, d'anciens résultats sont pris en compte et combinés avec les nouvelles observations. Cela peut éviter l'influence des données erronées dues à une erreur de mesure par exemple, en particulier dans le cas où le nombre des données est faible. Les lois non informatives sont classiquement choisies pour la loi *a priori*. Ce choix est souvent fait lorsqu'aucune information ne permet de mettre le poids sur une valeur plus qu'une autre pour le paramètre. Du point de vue facilité du calcul, les lois conjuguées sont souvent utilisées dans l'estimation Bayésienne. Une loi *a priori* est conjuguée pour la vraisemblance si la loi *a posteriori* est de la même classe que la loi *a priori*. Ce point n'est pas valable pour les données groupées, dans ce cas le calcul de la loi *a posteriori* a recours aux méthodes

d'approximation ou de simulation, telle que la méthode de Monte Carlo par la chaîne de Markov (MCMC). En terme de convergence, la performance de la méthode MCMC est satisfaisante pour un bon choix de valeur initiale et de distribution de proposition (Section 2.5.1). Lorsque la loi *a posteriori* du paramètre est obtenue, l'estimateur du paramètre peut être obtenu sous différentes fonctions de perte, par exemple la moyenne de la loi *a posteriori* sous fonction de perte d'erreur quadratique. Dans le contexte industriel d'une nouvelle conception, la loi peut être proposée par les experts afin d'assurer la qualité de l'estimation même quand il y a très peu de données disponibles.

Trois modèles de distribution ont été utilisés pour représenter la loi de la durée de vie, les lois continues Log-normale et Weibull, ainsi que la loi discrète multinomiale. Les lois continues sont souvent utilisées pour modéliser les durées de vie. Mais, étant donné le manque d'information dans les données groupées, la loi multinomiale qui décrit l'information des données d'une manière exacte est proposée. Toutefois, une loi continue est simple d'application dans l'industrie pour analyser, par exemple, la fiabilité, surtout lorsque les données de défaillances sont peu nombreuses pendant les essais. Dans la littérature, la loi de Log-normale et la loi de Weibull sont souvent utilisées pour décrire la distribution de la durée de vie des pièces mécaniques. D'après les experts de l'automobile, la loi Log-normale est plus adaptée que la loi de Weibull pour les essieux souples. Mais dans l'analyse de la fiabilité, le taux de défaillance d'une loi Log-normale diminue avec l'évolution du temps, ce qui n'est pas réaliste. Il est donc proposé de choisir les modèles en fonction de ce qui est à évaluer. La loi Log-normale est un bon choix pour évaluer la durée de vie, la loi de Weibull est plus réaliste pour analyser le taux de défaillance pour les pièces mécaniques.

Le second objectif de la thèse, développé dans le Chapitre 3, visait à prédire la variabilité de la durée de vie par la simulation numérique. Il a été supposé que cette variabilité est principalement liée à la courbe de Wöhler, représentant le comportement en fatigue d'un matériau. Les durées de vie caractérisées pour des éprouvettes ou des assemblages de tôles soudées ne mènent pas au même comportement, les données expérimentales Renault sont obtenues sur des structures. Une stratégie d'identification de la variabilité liée à la courbe de Wöhler a donc été proposée pour obtenir le comportement des cordons de soudure d'essieu souple. La stratégie a été de réussir à obtenir des dispersions comparables pour les durées de vie issues de la simulation par éléments finis et les durées de vie expérimentales. Des méthodes d'identification ont ainsi été développées avec une approche analytique et une approche simulation. Pour identifier la distribution des durées de vie numériques, l'approche simulation est basée sur un problème inverse avec minimisation d'une fonction coût mesurant la distance entre deux variables aléatoires. Dans la littérature, les moments probabilistes ou la distribution cumulative sont souvent utilisés comme fonction coût. Dans le travail réalisé, six fonctions coût ont été proposées, paramétrique ou non paramétrique à l'aide des statistiques. Parmi ces fonctions, la fonction coût basée sur le test statistique de Cramér-Von-Mises à un et à deux échantillons a montré les meilleures performances pour la précision de l'identification. La performance des méthodes proposées a été étudiée pour différents exemples numériques, il a été montré que globalement, la stratégie d'identification est robuste et efficace. Il est à noter que les fonctions coût proposées pour l'identification de la variabilité liée à la courbe de Wöhler peuvent être utilisées pour d'autres problèmes d'identification stochastiques.

Dans le cadre de l'approche simulation, la courbe de Wöhler est modélisée par la loi de Basquin, caractérisée par sa coordonnée et sa pente. Dans un premier temps, l'origine est considérée stochastique et la pente fixée. Dans un deuxième temps, la pente déterministe est considérée inconnue, l'optimisation multi-niveaux adaptée est proposée. Cette approche d'identification peut être appliquée séparément pour chaque cordon de soudure, et également pour plusieurs cordons de soudure en sommant leurs fonctions coût. S'il est souhaité une courbe de Wöhler pour tous les cordons de soudure d'un même type de pièce, cette dernière identification multi-sites serait plus intéressante. La méthodologie d'identification proposée dans ce travail de thèse est en effet adaptable en fonction du problème posé et des objectifs visés.

Afin d'évaluer la qualité de l'identification, plusieurs études ont été menées sur les sources d'erreur. D'une part, l'erreur peut provenir des données censurées de type groupé, par manque d'information de ces données, mais également de la méthode d'estimation. D'autre part, une erreur peut être induite par la méthodologie proposée, la construction des fonctions coût nécessite en effet des tirages d'échantillons représentant une loi de distribution. La taille de l'échantillon peut influencer la valeur des fonction coût, et donc le performance des algorithmes d'optimisation. Les études ont montré que, l'erreur liée à la méthodologie devient faible avec l'augmentation de la taille de l'échantillon. Par contre, l'erreur liée aux données groupées est assez significative, il a été observé que ce type de données induit une perte importante d'information. Pour une même loi, des données exactes et groupées mèneraient à une estimation bien différente. Pour cela, des études plus poussées restent à faire pour quantifier l'influence des données groupées, en termes de nombre d'intervalles et de largeur d'intervalles.

Plusieurs études ont été menées sur l'erreur d'identification. D'une part, l'erreur peut venir de l'estimation des données censurées, où il y a un manque d'information des données censurées et l'erreur de la méthode d'estimation. D'autre part, l'erreur existe dans la méthodologie, car pour construire les fonctions de coût il faut réaliser un tirage des échantillons suivant une loi de distribution. Plus précisément, la taille d'échantillon peut influencer la valeur des fonctions coûts et la performance des algorithmes d'optimisation. Dans les études proposées, l'erreur sur les résultats identifiés de la méthodologie liée à la taille d'échantillon devient faible avec l'augmentation de la taille d'échantillon. Par contre, l'erreur liée aux données groupées est assez remarquable. L'étude montre que le fait de grouper induit une forte réduction de l'information par rapport aux données exactes. L'estimation des données groupées pourrait donner un résultat bien différent que celui obtenu avec les données exactes pour une même loi au départ. Ce phénomène nécessite des études supplémentaires pour quantifier l'influence des données groupées, plus en détail, le nombre d'intervalles et de largeur d'intervalles.

Le dernier chapitre de ce manuscrit a été consacré à appliquer les stratégies d'identification sur les données de projets Renault. La méthodologie d'identification a conduit à l'obtention de courbes de Wöhler semblables pour les différentes méthodes, ce qui contribue à valider la performance de la méthodologie proposée. Une interface sous le logiciel MATLAB a été développée afin de faciliter l'utilisation de la méthodologie d'identification. De plus, pour une application industrielle, un calcul probabiliste de la durée de vie par éléments finis et un critère de corrélation essai-calcul ont été proposés.

Dans le contexte de validation des châssis automobiles en fatigue, des perspectives sont maintenant proposées, faisant suite aux développements sur le traitement des données expérimentales, et sur la prise en compte d'incertitudes dans la simulation numérique.

- Les méthodes proposées dans ce manuscrit pour estimer les données groupées expérimentales sont de type paramétrique. Une approche non-paramétrique est souhaitable, sans hypothèse sur la loi de distribution, ce qui pourrait être approprié pour estimer les données groupées.
- Le manque d'information dans les données groupées est un frein dans l'estimation de la durée de vie expérimentale. Il est envisagé de quantifier ce manque d'information par rapport aux données exactes dans le contexte d'essais Renault. Ceci pourrait permettre de quantifier le nombre d'essais nécessaires pour obtenir une loi de durée de vie associée à une erreur, soit encore garantir un niveau de confiance satisfaisant avec un nombre d'essais optimal.
- La méthodologie d'identification a été appliquée sur des projets Renault pour un modèle probabiliste de courbe de Wöhler avec un paramètre stochastique et une pente fixée sur chaque cordon de soudure pris individuellement. Il serait pertinent de réaliser l'identification sur plusieurs cordons de soudure afin d'avoir une loi à l'échelle de la structure. De plus, l'identification de la pente de la loi de Basquin permettrait d'obtenir une valeur calibrée adaptée pour le calcul de ce type de structure.
- Afin de quantifier plus précisément la variabilité liée à la courbe de Wöhler, il serait intéressant d'enrichir la base de données expérimentales. Les résultats obtenus permettraient de mieux maîtriser la variabilité pour un type de pièce. Ces informations seraient utiles à la prédiction de l'incertitude dans les zones d'intérêt des projets à venir, avec une réduction du nombre d'essais synthétiques sur les nouvelles structures. Cette approche probabiliste peut être étendue aux autres essais et à d'autres structures.

Annexe A

Calculation of Fisher information for grouped data

A.1 Fisher information of Weibull distribution

For a realisation $0 = t_0 < t_1 < \dots < t_{k-1} < t_k = \infty$, let us set :

$$\begin{aligned} P_j &= F(t_j; \theta) - F(t_{j-1}; \theta) = \left[1 - e^{-\left(\frac{t_j}{\alpha}\right)^\beta} \right] - \left[1 - e^{-\left(\frac{t_{j-1}}{\alpha}\right)^\beta} \right] \\ &= -e^{-\left(\frac{t_j}{\alpha}\right)^\beta} + e^{-\left(\frac{t_{j-1}}{\alpha}\right)^\beta} \end{aligned} \quad (\text{A.1})$$

The deviation with respect to α and β , for $1 < j < k$ are given as follows

$$\frac{\partial P_j}{\partial \alpha} = -\frac{\beta}{\alpha} \left[e^{-\left(\frac{t_j}{\alpha}\right)^\beta} \left(\frac{t_j}{\alpha}\right)^\beta + e^{-\left(\frac{t_{j-1}}{\alpha}\right)^\beta} \left(\frac{t_{j-1}}{\alpha}\right)^\beta \right] \quad (\text{A.2})$$

$$\frac{\partial P_j}{\partial \beta} = e^{-\left(\frac{t_j}{\alpha}\right)^\beta} \left(\frac{t_j}{\alpha}\right)^\beta \log\left(\frac{t_j}{\alpha}\right) + e^{-\left(\frac{t_{j-1}}{\alpha}\right)^\beta} \left(\frac{t_{j-1}}{\alpha}\right)^\beta \log\left(\frac{t_{j-1}}{\alpha}\right) \quad (\text{A.3})$$

$$\begin{aligned} \frac{\partial^2 P_j}{\partial \alpha^2} &= -\beta t_j^\beta e^{-\left(\frac{t_j}{\alpha}\right)^\beta} \left[\beta t_j^\beta \alpha^{-(2\beta+2)} - (\beta+1) \alpha^{-(\beta+2)} \right] \\ &\quad + \beta t_{j-1}^\beta e^{-\left(\frac{t_{j-1}}{\alpha}\right)^\beta} \left[\beta t_{j-1}^\beta \alpha^{-(2\beta+2)} - (\beta+1) \alpha^{-(\beta+2)} \right] \end{aligned} \quad (\text{A.4})$$

$$\begin{aligned} \frac{\partial^2 P_j}{\partial \beta^2} &= e^{-\left(\frac{t_j}{\alpha}\right)^\beta} \left(\frac{t_j}{\alpha}\right)^\beta \left[\log\left(\frac{t_j}{\alpha}\right) \right]^2 \left[-\left(\frac{t_j}{\alpha}\right)^\beta + 1 \right] \\ &\quad + e^{-\left(\frac{t_{j-1}}{\alpha}\right)^\beta} \left(\frac{t_{j-1}}{\alpha}\right)^\beta \left[\log\left(\frac{t_{j-1}}{\alpha}\right) \right]^2 \left[-\left(\frac{t_{j-1}}{\alpha}\right)^\beta + 1 \right] \end{aligned} \quad (\text{A.5})$$

For $i=1$, $P_1 = F(t_1; \theta)$

$$\frac{\partial P_1}{\partial \alpha} = -\frac{\beta}{\alpha} e^{-\left(\frac{t_1}{\alpha}\right)^\beta} \left(\frac{t_1}{\alpha}\right)^\beta \quad (\text{A.6})$$

$$\frac{\partial P_1}{\partial \beta} = \left(\frac{t_1}{\alpha}\right)^\beta \log\left(\frac{t_1}{\alpha}\right) e^{-\left(\frac{t_1}{\alpha}\right)^\beta} \quad (\text{A.7})$$

$$\frac{\partial^2 P_1}{\partial \alpha^2} = -\beta t_1^\beta e^{-\left(\frac{t_1}{\alpha}\right)^\beta} \left[\beta t_1^\beta \alpha^{-(2\beta+2)} - (\beta+1) \alpha^{-(\beta+2)} \right] \quad (\text{A.8})$$

$$\frac{\partial^2 P_1}{\partial \beta^2} = e^{-\left(\frac{t_1}{\alpha}\right)^\beta} \left(\frac{t_1}{\alpha}\right)^\beta \left[\log\left(\frac{t_1}{\alpha}\right)\right]^2 \left[-\left(\frac{t_1}{\alpha}\right)^\beta + 1\right] \quad (\text{A.9})$$

For $i=k$, $P_k = 1 - F(t_k; \theta)$

$$\frac{\partial P_k}{\partial \alpha} = \frac{\beta}{\alpha} e^{-\left(\frac{t_k}{\alpha}\right)^\beta} \left(\frac{t_k}{\alpha}\right)^\beta \quad (\text{A.10})$$

$$\frac{\partial P_k}{\partial \beta} = -\left(\frac{t_k}{\alpha}\right)^\beta \log\left(\frac{t_k}{\alpha}\right) e^{-\left(\frac{t_k}{\alpha}\right)^\beta} \quad (\text{A.11})$$

$$\frac{\partial^2 P_k}{\partial \alpha^2} = \beta t_k^\beta e^{-\left(\frac{t_k}{\alpha}\right)^\beta} \left[\beta t_k^\beta \alpha^{-(2\beta+2)} - (\beta+1) \alpha^{-(\beta+2)}\right] \quad (\text{A.12})$$

$$\frac{\partial^2 P_k}{\partial \beta^2} = -e^{-\left(\frac{t_k}{\alpha}\right)^\beta} \left(\frac{t_k}{\alpha}\right)^\beta \left[\log\left(\frac{t_k}{\alpha}\right)\right]^2 \left[-\left(\frac{t_k}{\alpha}\right)^\beta + 1\right] \quad (\text{A.13})$$

The expression of the matrix I is presented in the text.

A.2 Fisher information of Log-normal distribution

After doing a log transformation, the log fatigue life follows a Normal distribution, taking Equation (2.12), the cumulative density distribution of normal distribution $N(\mu, \sigma)$ is substituted.

$$P_j = F(t_j; \theta) - F(t_{j-1}; \theta) = \frac{1}{2} \operatorname{erf} \left[\frac{t_j - \mu}{\sigma \sqrt{2}} \right] - \frac{1}{2} \operatorname{erf} \left[\frac{t_{j-1} - \mu}{\sigma \sqrt{2}} \right] \quad (\text{A.14})$$

where $\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$ is the error function.

Therefore

$$P_j = \frac{1}{\sqrt{\pi}} \int_{\frac{t_{j-1}-\mu}{\sigma\sqrt{2}}}^{\frac{t_j-\mu}{\sigma\sqrt{2}}} e^{-\eta^2} d\eta \quad (\text{A.15})$$

The deviation with respect to μ and σ are given as follows :

$$\frac{\partial P_j}{\partial \mu} = -\frac{1}{\sigma \sqrt{2\pi}} \left[e^{-\frac{(t_j-\mu)^2}{2\sigma^2}} - e^{-\frac{(t_{j-1}-\mu)^2}{2\sigma^2}} \right] \quad (\text{A.16})$$

$$\frac{\partial P_j}{\partial \sigma} = -\frac{1}{\sqrt{2\pi}} \left[\frac{t_j - \mu}{\sigma^2} e^{-\frac{(t_j-\mu)^2}{2\sigma^2}} - \frac{t_{j-1} - \mu}{\sigma^2} e^{-\frac{(t_{j-1}-\mu)^2}{2\sigma^2}} \right] \quad (\text{A.17})$$

$$\frac{\partial^2 P_j}{\partial \mu^2} = -\frac{\mu}{\sqrt{2\pi}\sigma^3} \left[(t_j - \mu) e^{-\frac{(t_j-\mu)^2}{2\sigma^2}} - (t_{j-1} - \mu) e^{-\frac{(t_{j-1}-\mu)^2}{2\sigma^2}} \right] \quad (\text{A.18})$$

$$\frac{\partial^2 P_j}{\partial \sigma^2} = -\frac{1}{\sqrt{2\pi}} \left[e^{-\frac{(t_j-\mu)^2}{2\sigma^2}} \left(-\frac{2(t_j - \mu)}{\sigma^3} + \frac{(t_j - \mu)^3}{\sigma^5} \right) - e^{-\frac{(t_{j-1}-\mu)^2}{2\sigma^2}} \left(-\frac{2(t_{j-1} - \mu)}{\sigma^3} + \frac{(t_{j-1} - \mu)^3}{\sigma^5} \right) \right] \quad (\text{A.19})$$

For $i=1$, $P_1 = F(t_1; \theta)$

$$\frac{\partial P_1}{\partial \mu} = -\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(t_1-\mu)^2}{2\sigma^2}} \quad (\text{A.20})$$

$$\frac{\partial P_1}{\partial \sigma} = -\frac{t_1 - \mu}{\sigma^2\sqrt{2\pi}} e^{-\frac{(t_1-\mu)^2}{2\sigma^2}} \quad (\text{A.21})$$

$$\frac{\partial^2 P_1}{\partial \mu^2} = -\frac{t_1 - \mu}{\sigma^3\sqrt{2\pi}} e^{-\frac{(t_1-\mu)^2}{2\sigma^2}} \quad (\text{A.22})$$

$$\frac{\partial^2 P_1}{\partial \sigma^2} = -\frac{t_1 - \mu}{\sqrt{2\pi}\sigma^3} e^{-\frac{(t_1-\mu)^2}{2\sigma^2}} \left[-2 + \frac{(t_1 - \mu)^2}{\sigma^2} \right] \quad (\text{A.23})$$

For $i=k$, $P_k = 1 - F(t_k; \theta)$

$$\frac{\partial P_k}{\partial \mu} = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(t_k-\mu)^2}{2\sigma^2}} \quad (\text{A.24})$$

$$\frac{\partial P_k}{\partial \sigma} = \frac{t_k - \mu}{\sigma^2\sqrt{2\pi}} e^{-\frac{(t_k-\mu)^2}{2\sigma^2}} \quad (\text{A.25})$$

$$\frac{\partial^2 P_k}{\partial \mu^2} = \frac{t_k - \mu}{\sigma^3\sqrt{2\pi}} e^{-\frac{(t_k-\mu)^2}{2\sigma^2}} \quad (\text{A.26})$$

$$\frac{\partial^2 P_k}{\partial \sigma^2} = \frac{t_k - \mu}{\sqrt{2\pi}\sigma^3} e^{-\frac{(t_k-\mu)^2}{2\sigma^2}} \left[-2 + \frac{(t_k - \mu)^2}{\sigma^2} \right] \quad (\text{A.27})$$

Annexe B

Non-homogeneous grouped data

The non homogeneous raw grouped data used are as follows, where the T_i represents the inspection interval and N_i the vector of number of failures corresponding to the time interval.

$$T_1 = [0; 0.62; 1.06; 1.13; \infty];$$

$$T_2 = [0; 0.64; 1.03; 1.47; \infty];$$

$$T_3 = [0; 0.53; 0.94; 1.42; \infty];$$

$$T_4 = [0; 0.53; 1.00; \infty];$$

$$T_5 = [0; 0.53; 0.76; \infty];$$

$$T_6 = [0; 0.53; 0.82; \infty];$$

$$T_7 = [0; 0.53; 0.87; \infty];$$

$$T_8 = [0; 0.46; 1; 1.12; \infty];$$

$$T_9 = [0; 0.49; 1; \infty];$$

$$T_{10} = [0; 0.63; \infty];$$

$$T_{11} = [0; 0.5; 0.58; \infty];$$

$$T_{12} = [0; 0.5; 0.75; \infty];$$

$$T_{13} = [0; 0.5; 0.75; 0.99; \infty];$$

$$T_{14} = [0; 0.5; 0.75; \infty];$$

$$T_{15} = [0; 0.5; 0.75; 1; 1.25; 1.5; \infty];$$

$$T_{16} = [0; 0.5; 0.75; 1; 1.25; \infty];$$

$$N_1 = [0; 0; 0; 2];$$

$$N_2 = [0; 0; 0; 2];$$

$$N_3 = [0; 0; 0; 2];$$

$$N_4 = [0; 0; 4];$$

$$N_5 = [0; 0; 2];$$

$$N_6 = [0; 0; 2];$$

$$N_7 = [0; 2; 0];$$

$$N_8 = [0; 2; 0; 0];$$

$$N_9 = [0; 0; 2];$$

$$N_{10} = [1; 1];$$

$$N_{11} = [1; 0; 1];$$

$$N_{12} = [1; 0; 1];$$

$$N_{13} = [1; 0; 0; 1];$$

$$N_{14} = [1; 1; 0];$$

$$N_{15} = [0; 1; 0; 0; 0; 1];$$

$$N_{16} = [0; 0; 0; 0; 2];$$

Annexe C

Bayes estimation for Log-normal distribution of exact data with conjugate prior distribution

After a log transformation of the data, the estimation of Log-normal distribution could be considered as the estimation of normal distribution $X \sim \mathcal{N}(\mu, \sigma^2)$. The conjugate prior distribution of normal distribution for μ, σ is normal-inverse-gamma distribution. (reference)

Set $\tau = \frac{1}{\sigma^2}$ and suppose that $X \sim \mathcal{N}(\mu, \frac{1}{\tau})$, the parameter τ is said to be the precision. The conjugate prior distribution of μ, τ is thus normal-gamma distribution as follows

- $\tau \sim \text{Gamma}(\alpha, \beta)$, $\alpha > 0, \beta > 0$
- Conditionally on τ , $\mu \sim \mathcal{N}(\nu, \frac{1}{k\tau})$, $k > 0, \nu \in \mathbb{R}$

Thus the prior distribution is

$$\begin{aligned} \pi_{pr} &= \pi_{pr}(\mu|\tau)\pi_{pr}(\tau) \\ &= \frac{\beta^\alpha}{\Gamma(\alpha)} \tau^{\alpha-1} e^{-\beta\tau} \frac{1}{\sqrt{2\pi}} (k\tau)^{1/2} \exp\left[-\frac{k\tau}{2}(\mu - \nu)^2\right] \\ &\propto \tau^{\alpha-1/2} \exp\left[-\tau\left[\beta + \frac{k}{2}(\mu - \nu)^2\right]\right] \end{aligned} \quad (\text{C.1})$$

Then we have the observations of *r.v.* X, X_1, X_2, \dots, X_n , they are independently identically distributed from $\mathcal{N}(\mu, \frac{1}{\tau})$, so the likelihood function is given as follows

$$\mathcal{L}(\mu, \tau|x) = f(x|\mu, \tau) = \frac{\tau^{n/2}}{(2\pi)^{n/2}} \exp\left[-\frac{\tau}{2} \sum_{i=1}^n (x_i - \mu)^2\right]. \quad (\text{C.2})$$

According to the Bayes formulation in Equation (2.48), the posterior probability density function is

$$p(\tau, \mu|x) \propto \tau^{\alpha+\frac{n}{2}-\frac{1}{2}} \exp\left(-\tau\left[\beta + \frac{k}{2}(\mu - \nu)^2 + \frac{1}{2} \sum_{i=1}^n (x_i - \mu)^2\right]\right). \quad (\text{C.3})$$

After the calculation, the posterior distribution is

$$p(\tau, \mu|x) \propto \tau^{\alpha' + \frac{n}{2} - \frac{1}{2}} \exp\left(-\tau \left[\beta' + \frac{k'}{2}(\mu - \nu')^2\right]\right). \quad (\text{C.4})$$

with :

$$\begin{aligned} \alpha' &= \alpha + \frac{n}{2} \\ \beta' &= \beta + \frac{1}{2} \frac{nk}{n+k} (\bar{x} - \nu)^2 + \frac{1}{2} \sum (x_i - \bar{x})^2 \\ k' &= k + n \\ \nu' &= \frac{k\nu + n\bar{x}}{k+n} \end{aligned}$$

Annexe D

Calcul de la statistique de test de Cramér-von Mises

La statistique de CVM est donnée par :

$$\omega_n^2 = \int_{-\infty}^{+\infty} [F_n^*(x) - F_0(x)]^2 dF(x), \quad (\text{D.1})$$

où F_n^* est la fonction de densité cumulative empirique de la variable aléatoire x et F_0 la fonction cumulative théorique.

La forme de calcul de cette intégrale est [89] :

$$\omega_n^2 = \frac{1}{12n} + \sum_{i=1}^n \left(u_{(i)} - \frac{2i-1}{2n} \right)^2, \quad (\text{D.2})$$

en supposant que $x_{(n)}$ est la statistique d'ordre de l'observation x , et $u = F_0(x)$ de sorte que $u_{(n)} = F_0(x_{(n)})$.

Démonstration. Changer la variable de t à u par $u = F_0(x)$, donc $x = F_0^{-1}(u)$ dans l'intégrale et la représenter par une fonction indicatrice :

$$\begin{aligned} \omega_n^2 &= n \int (F_n^*(x) - F_0(x))^2 dF_0(x) \\ &= n \int (F_n^*(F_0^{-1}(u)) - u)^2 du \\ &= n \int_0^1 \left(\frac{1}{n} \sum_{i=0}^n \mathbb{1}_{\{F_0(X_i) \leq u\}} - u \right)^2 du. \end{aligned} \quad (\text{D.3})$$

Notez que,

$$\mathbb{1}_{\{F_0(X_{(i)}) \leq u\}} \stackrel{d}{=} \mathbb{1}_{\{U_{(i)} \leq u\}}. \quad (\text{D.4})$$

L'intégrale devient

$$\begin{aligned} \omega_n^2 &= n \int_0^1 \left(\frac{1}{n} \sum_{i=0}^n \mathbb{1}_{\{F_0(X_i) \leq u\}} - u \right)^2 du \\ &= n \sum_{i=0}^n \int_{U_{(i)}}^{U_{(i+1)}} \left(\frac{i}{n} - u \right)^2 du, \end{aligned} \quad (\text{D.5})$$

où $U_{(0)} = 0, U_{(n+1)} = 1$.

En résolvant l'intégrale définie ci-dessus, on obtient :

$$\begin{aligned}\omega_n^2 &= n \sum_{i=0}^n \frac{1}{3} \left(\left(U_{(i+1)} - \frac{i}{n} \right)^3 - \left(U_{(i)} - \frac{i}{n} \right)^3 \right) \\ &= n \sum_{i=1}^n \frac{1}{3} \left(\left(U_{(i)} - \frac{i-1}{n} \right)^3 - \left(U_{(i)} - \frac{i}{n} \right)^3 \right)\end{aligned}\tag{D.6}$$

On note $y_i = U_{(i)} - \frac{2i-1}{2n}$, et on a

$$\begin{aligned}U_{(i)} - \frac{i-1}{n} &= y_i + \frac{1}{2n}, \\ U_{(i)} - \frac{i}{n} &= y_i - \frac{1}{2n}.\end{aligned}\tag{D.7}$$

Selon le théorème binomial $a^3 - b^3 = (a-b)(a^2 + ab + b^2)$.

$$\begin{aligned}\omega_n^2 &= n \sum_{i=0}^n \frac{1}{3} \left(\left(y_i + \frac{1}{2n} \right)^3 - \left(y_i - \frac{1}{2n} \right)^3 \right) \\ &= n \sum_{i=0}^n \frac{1}{3} \left(\left(y_i + \frac{1}{2n} \right)^2 + \left(y_i + \frac{1}{2n} \right) \left(y_i - \frac{1}{2n} \right) + \left(y_i - \frac{1}{2n} \right)^2 \right) \\ &\quad \left(y_i + \frac{1}{2n} - y_i + \frac{1}{2n} \right) \\ &= n \sum_{i=0}^n \frac{1}{3} \left(3y_i^2 + \frac{1}{4n^2} \right) \\ &= \frac{1}{12n} + \sum_{i=1}^n y_i^2 \\ &= \frac{1}{12n} + \sum_{i=1}^n \left[\left(U_{(i)} - \frac{2i-1}{2n} \right) \right]^2.\end{aligned}\tag{D.8}$$

□

Nous obtenons donc le résultat souhaité.

Annexe E

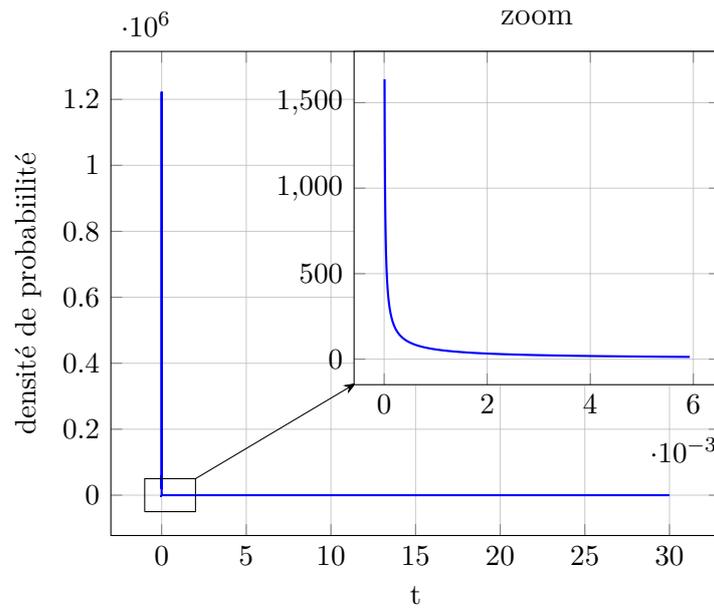
Vérification de la densité de probabilité

Nous supposons que les variables aléatoires a et b suivent toutes les deux une loi Log-normale, $a \sim \mathcal{LN}(\mu_a, \sigma_a)$ et $b \sim \mathcal{LN}(\mu_b, \sigma_b)$. Nous avons pris les valeurs $\mu_a = 32, \sigma_a = 2, \mu_b = 0.6, \sigma_b = 0.15$ pour l'exemple numérique suivant. Avec ces lois, l'intégrale de l'Équation (3.5) est calculée numériquement. Le calcul de l'intégrale de f_T sur \mathbb{R}^+ est fait en deux étapes, premièrement l'intégrale par rapport à v est estimé pour un t fixé, puis l'intégrale par rapport à t sur $[0, +\infty)$ est estimé. Même si le domaine d'intégration est sur $[0, +\infty)$, la fonction égale quasiment à 0 pour $t \geq 30$, donc on a décidé de calculer l'intégrale sur $[0, 30]$.

Trois méthodes numériques pour calculer l'intégrale sont étudiées, la méthode de Simpson simple, la méthode de Simpson 3/8 et la méthode de Boole, toutes ces méthodes visent à discrétiser le domaine d'intégration avec un pas choisi, puis d'approcher l'intégrale comme la somme pondérée par des coefficients d'aire de sous-domaines approchés. Parmi les méthodes, la méthode de Boole donne une meilleure précision d'après les résultats obtenus. Il faut remarquer que la précision de toutes ces méthodes est dépendante de la taille du pas choisie.

Suite à différents essais, nous avons choisi le pas de v égale à 10^{-6} pour un t fixé, et pour t le pas est choisi différemment pour les intervalles de t , soit 10^{-11} pour $t \in [0, 10^{-8}]$, 10^{-10} pour $t \in [10^{-8}, 10^{-7}]$, 10^{-9} pour $t \in [10^{-7}, 10^{-6}]$... 10^{-2} pour $t \in [1, 30]$. L'estimation de l'intégrale est égale à 0.9447 qui donne une erreur de 5.6% par rapport à la valeur cible de 1. Avec le pas choisi pour t , la densité de probabilité de la durée de vie est présentée Figure E.1 pour le jeu de paramètre choisi.

Il est vérifié que l'Équation (3.5) donne une densité de probabilité de la variable aléatoire T . Cependant la courbe illustrée (Figure E.1) ne ressemble pas à ce que nous avons obtenu avec les vraies mesures expérimentales (voir Figure 2.4). Cela veut dire que soit la type de loi (loi Log-normale) n'est pas bien choisi pour les variables aléatoires a et b , soit les paramètres de la loi Log-normale ne sont pas bien choisis. Des études complémentaires sont nécessaires à son estimation.

FIGURE E.1 – Densité de probabilité du T

Annexe F

Interface d'identification de la variabilité liée à la courbe de Whöler

Afin de pouvoir appliquer les méthodologies d'identification proposées sur les futurs projets Renault, une interface sous MATLAB a été développée pour faciliter l'utilisation des travaux de thèse. Les entrées de l'interface sont des fichiers Excel. Une initialisation avec l'étape 0 est nécessaire pour préparer les séquences amplitudes des contraintes et nombres de cycles correspondant, issues du comptage Rainflow permettant de compter les cycles à partir de l'enregistrement temporel des contraintes. L'interface est composée de 2 étapes numérotées 1 et 2, respectivement représentatives des développements des chapitres 3 et 4, soient l'estimation des données expérimentales et l'identification de la des variabilités liées à la courbe de Wöhler.

Étape 0 - Comptage Rainflow pour une forme temporelle des contraintes

Suite au calcul par éléments finis, les contraintes sous forme temporelle sont extraites du solveur NASTRAN et comptées par la méthode Rainflow pour obtenir les amplitudes de contraintes et les nombres de cycle correspondants. Ces données seront utilisées dans l'étape 2 suivante. Un script est proposé pour réaliser cette démarche. Les contraintes pour les éléments finis d'intérêt sont à importer au lancement du script. Les amplitudes de contraintes et les nombres de cycles correspondants sont écrits dans un fichier Excel en sortie.

Étape 1 - Estimation des données expérimentales

Les données brutes des observations expérimentales sont chargées dans un fichier Excel. La feuille Excel (Figure F.1) est nommée par le nom du projet et notée '`_sym`' pour symétrie ou '`_nsym`' pour non symétrie. Pour chaque essai, le temps de ressuage est indiqué en pourcentage par rapport à une durée de référence, soit en ordre croissant ou ordre décroissant. Pour chaque moment de ressuage, la longueur des fissures est également indiquée si elle est disponible. Le programme MATLAB lit le temps de ressuage et la longueur de fissure pour créer les données groupées et faire l'estimation de la loi de durée

Numéro des essais	Temps de ressuage	Traverse	Traverse	Traverse	Traverse	Traverse	Traverse	Coupelle	Coupelle	Tête essieu	Tête essieu
	%										
1	150%										
2	150%										
3	150%										
4	150%										
5	150%										
6	103%										
7	150%										
8	150%										
9	150%										

FIGURE F.1 – Feuille Excel des données expérimentales

de vie. Cette étape 1 d'interface est présentée par la Figure F.2. L'affichage de la densité de probabilité estimée de chaque cordon de soudure est à choisir. Le nom du fichier de sortie est aussi demandé.

Main_Estim_GUI

Input

Input file : C:\Users\la019694\OneDrive - Alliance\Documents\MATLA Select

Sheetname : X98 B4_exemple_sym

Display figures? Yes

Output

Output file : C:\Users\la019694\OneDrive - Alliance\Documents\MATLA Select

Go

FIGURE F.2 – Etape 1 de l'interface

A la sortie de l'interface, les paramètres de la loi Log-normale (μ_T, σ_T), 10^{me} percentile et la médiane sont écrits dans un fichier Excel (Figure F.3). La densité de probabilité de la durée de vie pour chaque cordon de soudure est affichée (Figure 4.5) si l'option est 'Yes'. Les données brutes du nombre de défaillances dans chaque intervalle sont affichées et normalisées sous la forme d'histogramme. La valeur des bornes inférieures est utilisée pour représenter la durée de vie de chaque intervalle.

	A	B	C	D	E	F
1		mu	sigma	B10	B50	
2						
3	Numéro des cordons					
4			Valeur des paramètres			
5						
6						
7						
8						
9						
10						
11						
12						
	X52_ph1_sym	X52_ph1_nsym	X52_ph2_sym	X52_ph2_nsym		

FIGURE F.3 – Paramètres estimés à partir des données expérimentales

Étape 2 - Identification de la variabilité liée à la courbe de Wöhler

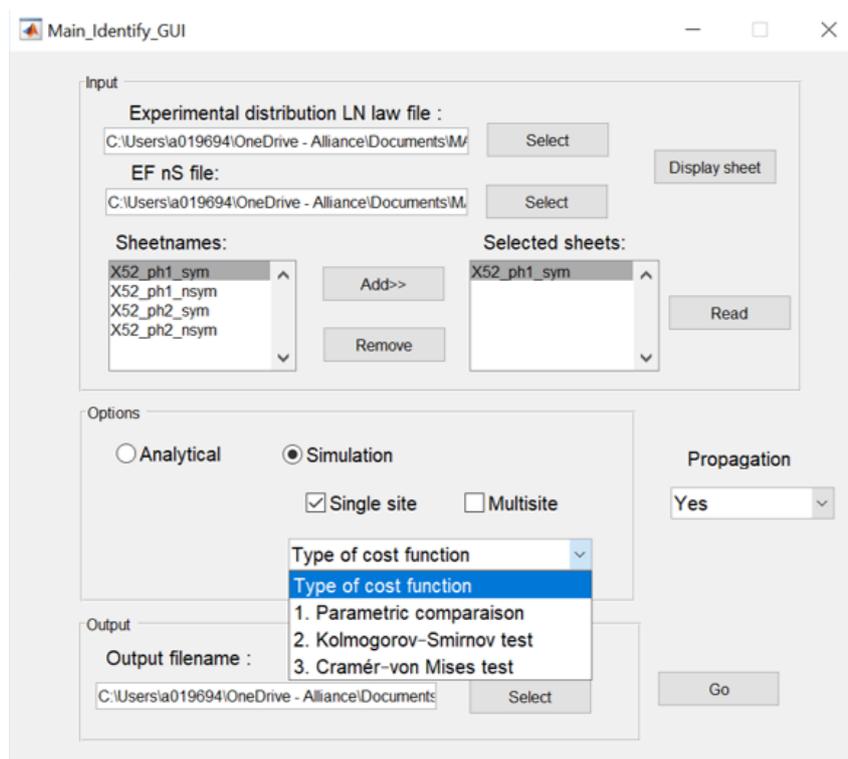


FIGURE F.4 – Etape 2 de l'interface

L'étape 2 est présentée dans la Figure F.4. En entrée, d'un point de vue expérimental, la loi Log-normale estimée de la durée de vie pour chaque cordon de soudure obtenue à l'issue de l'étape 1 est demandée (Fichier dans la Figure F.3). De plus, une colonne associée au poids de chaque cordon de soudure est ajoutée (0 ou 1), servant à décider si ce cordon est pris en compte ou pas dans l'optimisation multi-zones. D'un point de vue simulation numérique, les amplitudes de contraintes et les nombres de cycles correspondant obtenus avec la technique de comptage Rainflow en étape 0 sont demandés. Les données des différents projets sont écrites dans des feuilles Excel différentes. Les noms des feuilles

sont affichés dans l'interface. L'interface permet de choisir soit l'approche analytique soit une approche simulation des méthodologies. L'identification des cordons de soudure peut être réalisée pour un seul cordon d'une zone ou pour plusieurs cordons (multi-zones). Les paramètres identifiés de la courbe de Wöhler (μ_a, σ_a) d'un-site ou multi-zones sont écrits dans un fichier Excel en sortie (Figure F.5), également le temps d'optimisation et le poids pour chaque cordon de soudure.

	A	B	C	D	E	F
1	name	mu_a	sigma_a	t(s)	Weight	
2				125.832369	1	
3	Numéro			179.899465	1	
4	des		Valeurs des	123.811772	1	
5	cordons		paramètres	138.628045	1	
6				118.377055	1	
7	multisite			401.323345		
8	Parametric comparaison					
9						
10						
11						
12						

FIGURE F.5 – Fichier de résultat

Bibliographie

- [1] J. Schijve, “Fatigue of structures and materials in the 20th century and the state of the art,” *International journal of fatigue*, vol. 25, no. 8, pp. 679–702, 2003.
- [2] J. Schijve, “Statistical distribution functions and fatigue of structures,” *International Journal of Fatigue*, vol. 27, no. 9, pp. 1031–1039, 2005.
- [3] M. Zoroufi and A. Fatemi, “Experimental durability assessment and life prediction of vehicle suspension components : a case study of steering knuckles,” *Proceedings of the Institution of Mechanical Engineers, Part D : Journal of Automobile Engineering*, vol. 220, no. 11, pp. 1565–1579, 2006.
- [4] C. M. Sonsino, H. Kaufmann, J. Foth, and F. Jauch, “Fatigue strength of driving shafts of automatic transmission gearboxes under operational torques,” *SAE transactions*, pp. 635–648, 1997.
- [5] Y. L. Lee, J. Pan, R. Hathaway, and M. Barkey, *Fatigue testing and analysis : theory and practice*, vol. 13. Butterworth-Heinemann, 2005.
- [6] Y. Murakami, “Material defects as the basis of fatigue design,” *International Journal of Fatigue*, vol. 41, pp. 2–10, 2012.
- [7] Y. L. Lee and D. Taylor, “Stress-based fatigue analysis and design,” in *Fatigue testing and analysis (theory and practice)*, ch. 4, pp. 103–180, Oxford, UK : Elsevier Butterworth-Heinemann, 2005.
- [8] J. Chiquet, N. Limnios, and M. Eid, “Piecewise deterministic Markov processes applied to fatigue crack growth modelling,” *Journal of Statistical Planning and Inference*, vol. 139, no. 5, pp. 1657–1667, 2009.
- [9] X. Long, K. Liu, C. Jiang, Y. Xiao, and S. Wu, “Uncertainty propagation method for probabilistic fatigue crack growth life prediction,” *Theoretical and Applied Fracture Mechanics*, vol. 103, p. 102268, 2019.
- [10] J. Thomas, G. Perroud, A. Bignonnet, and D. Monnet, “Fatigue design and reliability in the automotive industry,” in *European Structural Integrity Society*, vol. 23, pp. 1–11, Elsevier, 1999.
- [11] M. A. Laribi, R. TieBi, S. Tamboura, M. Shirinbayan, A. Tcharkhtchi, H. B. Dali, and J. Fitoussi, “Sheet molding compound automotive component reliability using a micromechanical damage approach,” *Applied Composite Materials*, vol. 27, no. 5, pp. 693–715, 2020.

- [12] P. Beaumont, *Optimisation des plans d'essais accélérés Application à la tenue en fatigue de pièces métalliques de liaison au sol*. PhD thesis, Université d'Angers, 2013.
- [13] L. Godefroid, G. Faria, L. Câendido, and S. Araujo, "Fatigue failure of a welded automotive component," *Procedia Materials Science*, vol. 3, pp. 1902–1907, 2014. 20th European Conference on Fracture.
- [14] W. Fricke, "Fatigue analysis of welded joints : state of development," *Marine structures*, vol. 16, no. 3, pp. 185–200, 2003.
- [15] G. Schueller, "On the treatment of uncertainties in structural mechanics and analysis," *Computers & Structures*, vol. 85, no. 5-6, pp. 235–243, 2007.
- [16] K. A. Zakaria, S. Abdullah, and M. Ghazali, "A review of the loading sequence effects on the fatigue life behaviour of metallic materials.," *Journal of Engineering Science & Technology Review*, vol. 9, no. 5, 2016.
- [17] W. Fauriat, *Stochastic modeling of road-induced loads for reliability assessment of chassis and vehicle components through simulation*. PhD thesis, Université Blaise Pascal-Clermont-Ferrand II, 2016.
- [18] É. Charkaluk, *Fatigue des matériaux métalliques : quelques contributions à une approche dissipative*. PhD thesis, Université de Lille, 2006.
- [19] G. Marsh, C. Wignall, P. R. Thies, N. Barltrop, A. Incecik, V. Venugopal, and L. Johannig, "Review and application of rainflow residue processing techniques for accurate fatigue damage estimation," *International Journal of Fatigue*, vol. 82, pp. 757–765, 2016.
- [20] C. Amzallag, J. Gerey, J. L. Robert, and J. Bahuaud, "Standardization of the rain-flow counting method for fatigue analysis," *International journal of fatigue*, vol. 16, no. 4, pp. 287–293, 1994.
- [21] ASTM, "Standard practices for cycle counting in fatigue analysis," tech. rep., American Society for Testing and Material, 1997.
- [22] Q. Bader and E. Kadum, "Mean stress correction effects on the fatigue life behavior of steel alloys by using stress life approach theories," *International Journal of Engineering & Technology IJETIJENS*, vol. 10, p. p50, 2014.
- [23] A. Bhaduri, *Mechanical properties and working of metals and alloys*, vol. 264, ch. Fatigue. Springer, 2018.
- [24] L. Zhao, H. Cai, T. Wang, and S. Zheng, "Durability assessment of automotive structures under random variable amplitude loading," *Advances in Mechanical Engineering*, vol. 10, no. 4, p. 1687814018771766, 2018.
- [25] B. Weber, *Fatigue multiaxiale des structures industrielles sous chargement quelconque*. PhD thesis, Lyon, INSA, 1999.
- [26] F. Conle and C.-C. Chu, "Fatigue analysis and the local stress–strain approach in complex vehicular structures," *International journal of fatigue*, vol. 19, no. 93, pp. 317–323, 1997.

-
- [27] J. F. Barbosa, J. A. Correia, R. Freire Junior, S.-P. Zhu, and A. M. De Jesus, "Probabilistic SN fields based on statistical distributions applied to metallic and composite materials : State of the art," *Advances in Mechanical Engineering*, vol. 11, no. 8, p. 1687814019870395, 2019.
- [28] P. H. Wirsching, "Statistical summaries of fatigue data for design purposes," tech. rep., NASA, 1983.
- [29] F. Goedel, G. P. Mezzomo, and Z. M. C. Pravia, "Fatigue lifespan of a fillet welded joint-hybrid approach to obtain the SN curve with a reduced number of tests," *Latin American Journal of Solids and Structures*, vol. 15, no. 10, 2018.
- [30] H. Sutherland and P. Veers, "The development of confidence limits for fatigue strength data," in *2000 ASME Wind Energy Symposium*, p. 63, 2000.
- [31] S. Hanaki, M. Yamashita, H. Uchida, and M. Zako, "On stochastic evaluation of S-N data based on fatigue strength distribution," *International Journal of Fatigue*, vol. 32, no. 3, pp. 605–609, 2010.
- [32] C. Li, S. Wu, J. Zhang, L. Xie, and Y. Zhang, "Determination of the fatigue P–S–N curves—a critical review and improved backward statistical inference method," *International Journal of Fatigue*, vol. 139, p. 105789, 2020.
- [33] R. C. S. F. Júnior and A. S. Belísio, "Probabilistic S–N curves using exponential and power laws equations," *Composites Part B : Engineering*, vol. 56, pp. 582–590, 2014.
- [34] L. D'Angelo and A. Nussbaumer, "Estimation of fatigue SN curves of welded joints using advanced probabilistic approach," *International Journal of Fatigue*, vol. 97, pp. 98–113, 2017.
- [35] J. Klemenc and M. Fajdiga, "Estimating S–N curves and their scatter using a differential ant-stigmergy algorithm," *International journal of fatigue*, vol. 43, pp. 90–97, 2012.
- [36] D. G. Pavlou, "The theory of the sn fatigue damage envelope : generalization of linear, double-linear, and non-linear fatigue damage models," *International Journal of Fatigue*, vol. 110, pp. 204–214, 2018.
- [37] P. Strzelecki and T. Tomaszewski, "Application of Weibull distribution to describe SN curve with using small number specimens," in *AIP Conference Proceedings*, vol. 1780, p. 020007, AIP Publishing LLC, 2016.
- [38] ASTM, "Standard practices for statistical analysis of linear or linearized stress-life S-N and strain-life (ϵ -N) fatigue data," tech. rep., American Society for Testing and Material, 2015.
- [39] E. 1993-1-9, "Eurocode 3 : Design of steel structures-part 1-9 : Fatigue," *European Committee for Standardization : Brussels, Belgium*, 2005.
- [40] T. Sakai and A. Sugeta, "Publication of the second edition of standard evaluation method of fatigue reliability for metallic materials [standard regression method of S-N curves]," *Journal of the Society of Materials Science, Japan*, vol. 54, no. 1, pp. 37–43, 2005.

- [41] S. Petinoy, H. Reemsnyder, and A. Thayamballi, “The similitude of fatigue damage principle : application in SN curves-based fatigue design,” in *European Structural Integrity Society*, vol. 23, pp. 219–228, Elsevier, 1999.
- [42] S. C. Wu, Y. X. Liu, C. H. Li, G. Z. Kang, and S. L. Liang, “On the fatigue performance and residual life of intercity railway axles with inside axle boxes,” *Engineering Fracture Mechanics*, vol. 197, pp. 176–191, 2018.
- [43] S. Dekker, G. Wurzel, and R. Alderliesten, “A Bayesian tolerance interval estimation method for fatigue strength substantiation of rotorcraft dynamic components,” *International Journal of Fatigue*, vol. 92, pp. 333–344, 2016.
- [44] H. Nishino and K. Kakamu, “Grouped data estimation and testing of Gini coefficients using lognormal distributions,” *Sankhya B*, vol. 73, no. 2, pp. 193–210, 2011.
- [45] H. S. Migdadi and M. S. Al-Batah, “Bayesian inference based on the interval grouped data from the Weibull model with application,” *Journal of Advances in Mathematics and Computer Science*, pp. 1170–1183, 2014.
- [46] C. B. Guure, N. A. Ibrahim, and M. B. Adam, “Bayesian inference of the Weibull model based on interval-censored survival data,” *Computational and Mathematical Methods in Medicine*, vol. 2013, 2013.
- [47] A. Sadek and F. Alharbi, “Weibull-Bayesian analysis based on ranked set sampling,” *International Journal of Advanced Statistics and Probability*, vol. 2, no. 2, pp. 114–123, 2014.
- [48] D. Kundu and D. Mitra, “Bayesian inference of Weibull distribution based on left truncated and right censored data,” *Computational Statistics & Data Analysis*, vol. 99, pp. 38–50, 2016.
- [49] A. Ly, M. Marsman, J. Verhagen, R. Grasman, and E.-J. Wagenmakers, “A tutorial on Fisher information,” *Journal of Mathematical Psychology*, vol. 80, pp. 40–55, 2017.
- [50] K. Aludaat, M. Alodat, and T. Alodat, “Parameter estimation of Burr type X distribution for grouped data,” *Applied Mathematical Sciences*, vol. 2, no. 9, pp. 415–423, 2008.
- [51] J. Xia, J. Mi, and Y. Zhou, “On the existence and uniqueness of the maximum likelihood estimators of normal and lognormal population parameters with grouped data,” *Journal of Probability and Statistics*, 2009.
- [52] A. Geraci, F. Katki, L. McMonegal, B. Meyer, J. Lane, P. Wilson, J. Radatz, M. Yee, H. Porteous, and F. Springsteel, *IEEE standard computer dictionary : Compilation of IEEE standard computer glossaries*. IEEE Press, 1991.
- [53] S. Wang, M. Tomovic, and H. Liu, “Comprehensive reliability design of aircraft hydraulic system,” *Commercial Aircraft Hydraulic Systems*, pp. 115–169, 2016.
- [54] M. C. Bryson, “Heavy-tailed distributions : properties and tests,” *Technometrics*, vol. 16, no. 1, pp. 61–68, 1974.
- [55] S. Nadarajah, “Bathtub-shaped failure rate functions,” *Quality & Quantity*, vol. 43, no. 5, pp. 855–863, 2009.

-
- [56] M. K. Shakhathreh, A. J. Lemonte, and G. Moreno - Arenas, "The Log-normal modified Weibull distribution and its reliability implications," *Reliability Engineering & System Safety*, vol. 188, pp. 6–22, Aug. 2019.
- [57] W. J. Conover, "A Kolmogorov goodness-of-fit test for discontinuous distributions," *Journal of the American Statistical Association*, vol. 67, no. 339, pp. 591–596, 1972.
- [58] A. N. Pettitt and M. A. Stephens, "The Kolmogorov-Smirnov goodness-of-fit statistic with discrete and grouped data," *Technometrics*, vol. 19, no. 2, pp. 205–210, 1977.
- [59] S. Darmosiswoyo, "Kolmogorov-Smirnov test for grouped data," Master's thesis, Naval Postgraduate School, 1975.
- [60] B. Y. Lemeshko, E. V. Chimitova, and S. S. Kolesnikov, "Nonparametric goodness-of-fit tests for discrete, grouped or censored data," in *Proc. ASMDA 2007 Int. Conf*, pp. 1–7, 2007.
- [61] A. O'Hagan, "Bayesian statistics : principles and benefits," *Frontis*, pp. 31–45, 2004.
- [62] J. K. Ghosh, M. Delampady, and T. Samanta, *An introduction to Bayesian analysis : theory and methods*. Springer Science & Business Media, 2007.
- [63] H. Rue, S. Martino, and N. Chopin, "Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations," *Journal of the royal statistical society : Series b (statistical methodology)*, vol. 71, no. 2, pp. 319–392, 2009.
- [64] R. E. Kass, B. P. Carlin, A. Gelman, and R. M. Neal, "Markov chain Monte Carlo in practice : a roundtable discussion," *The American Statistician*, vol. 52, no. 2, pp. 93–100, 1998.
- [65] G. O. Roberts and J. S. Rosenthal, "General state space Markov chains and MCMC algorithms," *Probability Surveys*, vol. 1, pp. 20–71, 2004.
- [66] M. Evans and T. Swartz, "Methods for approximating integrals in statistics with special emphasis on Bayesian integration problems," *Statist. Sci.*, vol. 10, pp. 254–272, 08 1995.
- [67] V. Girardin and N. Limnios, *Applied probability*. Springer, 2018.
- [68] W. L. Martinez and A. R. Martinez, *Computational statistics handbook with MATLAB*, vol. 22. CRC press, 2015.
- [69] C. Geyer, "Practical Markov chain Monte Carlo," *Statistical Science*, vol. 7, pp. 473–483, Nov. 1992.
- [70] A. B. Owen, *Monte Carlo theory, methods and examples*. 2013.
- [71] C. Sherlock, P. Fearnhead, and G. O. Roberts, "The random walk metropolis : linking theory and practice through a case study," *Statistical Science*, vol. 25, no. 2, pp. 172–190, 2010.
- [72] R. M. Soland, "Bayesian analysis of the Weibull process with unknown scale and shape parameters," *IEEE Transactions on Reliability*, vol. 18, no. 4, pp. 181–184, 1969.

- [73] N. Bousquet, “A Bayesian analysis of industrial lifetime data with Weibull distributions,” tech. rep., INRIA, 2006.
- [74] J. S. Rosenthal, “Optimal proposal distributions and adaptive MCMC,” 2010.
- [75] T. Takaishi, “Markov chain Monte Carlo versus importance sampling in Bayesian inference of the GARCH model,” *Procedia Computer Science*, vol. 22, pp. 1056–1064, 2013.
- [76] K. P. Murphy, “Conjugate Bayesian analysis of the Gaussian distribution,” *def*, vol. 1, no. $2\sigma^2$, p. 16, 2007.
- [77] T. P. Minka, “Bayesian inference, entropy, and the multinomial distribution,” *Online tutorial*, 2003.
- [78] J. O. Berger, J. M. Bernardo, and D. Sun, “Overall Objective Priors,” *Bayesian Analysis*, vol. 10, no. 1, pp. 189 – 221, 2015.
- [79] B. V. Rosić, A. Kučerová, J. Šỳkora, O. Pajonk, A. Litvinenko, and H. G. Matthies, “Parameter identification in a probabilistic setting,” *Engineering Structures*, vol. 50, pp. 179–196, 2013.
- [80] N. Zabaras and B. Ganapathysubramanian, “A scalable framework for the solution of stochastic inverse problems using a sparse grid collocation approach,” *Journal of Computational Physics*, vol. 227, no. 9, pp. 4697–4735, 2008.
- [81] M. Ponski and A. Sluzalec, “Modeling and simulation of stochastic inverse problems in viscoplasticity,” *Transactions of the Indian Institute of Metals*, vol. 72, no. 10, pp. 2803–2817, 2019.
- [82] R. Rikards, A. Chate, and G. Gailis, “Identification of elastic properties of laminates based on experiment design,” *International Journal of Solids and Structures*, vol. 38, no. 30-31, pp. 5097–5115, 2001.
- [83] E. Zhang, J. Antoni, and P. Feissel, “Bayesian force reconstruction with an uncertain model,” *Journal of Sound and Vibration*, vol. 331, no. 4, pp. 798–814, 2012.
- [84] C. Mares, J. Mottershead, and M. Friswell, “Stochastic model updating : Part 1: Theory and simulated example,” *Mechanical Systems and Signal Processing*, vol. 20, no. 7, pp. 1674–1695, 2006.
- [85] G. Mariethoz, P. Renard, and J. Caers, “Bayesian inverse problem and optimization with iterative spatial resampling,” *Water Resources Research*, vol. 46, no. 11, 2010.
- [86] A. Tran and T. Wildey, “Solving stochastic inverse problems for property–structure linkages using data-consistent inversion and machine learning,” *JOM*, vol. 73, no. 1, pp. 72–89, 2021.
- [87] J. Latz, “On the well-posedness of Bayesian inverse problems,” *SIAM/ASA Journal on Uncertainty Quantification*, vol. 8, no. 1, pp. 451–482, 2020.
- [88] G. Saporta, *Probabilités, analyse des données et statistique*. éditions Technip, 2006.
- [89] T. W. Anderson, “On the distribution of the two-sample Cramér-von Mises criterion,” *The Annals of Mathematical Statistics*, pp. 1148–1159, 1962.

-
- [90] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proceedings of ICNN'95-international conference on neural networks*, vol. 4, pp. 1942–1948, IEEE, 1995.
- [91] D. Wang and L. Meng, "Performance analysis and parameter selection of PSO algorithms," *Acta Automatica Sinica*, vol. 42, no. 10, pp. 1552–1561, 2016.
- [92] K. M. Ramachandran and C. P. Tsokos, *Chapter 12 - Nonparametric Statistics*, pp. 491–530. Academic Press, Jan. 2021.
- [93] R. Thompson, "Bias of the one-sample Cramér-von Mises test," *Journal of the American Statistical Association*, vol. 61, no. 313, pp. 246–247, 1966.
- [94] F. Massey, "A note on the power of a non-parametric test," *The Annals of Mathematical Statistics*, vol. 21, no. 3, pp. 440–443, 1950.
- [95] C. Lemieux, *Monte Carlo and quasi-Monte Carlo sampling*. Springer Science & Business Media, 2009.
- [96] J. C. Helton and F. J. Davis, "Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems," *Reliability Engineering & System Safety*, vol. 81, no. 1, pp. 23–69, 2003.
- [97] E. Viennet, "Réseaux à fonctions de base radiales," in *Apprentissage connexionniste* (Y. Bennani, ed.), I2C Hermès, p. 105, Lavoisier, 2006. 18 pages.
- [98] J. Moody and C. J. Darken, "Fast learning in networks of locally-tuned processing units," *Neural Computation*, vol. 1, no. 2, pp. 281–294, 1989.
- [99] M. J. L. Orr *et al.*, "Introduction to radial basis function networks," 1996.
- [100] N. Benoudjit, C. Archambeau, A. Lendasse, J. A. Lee, and M. Verleysen, "Width optimization of the Gaussian kernels in radial basis function networks.," in *ESANN*, vol. 2, pp. 425–432, 2002.
- [101] M. Lázaro, I. Santamaría, and C. Pantaleón, "A new EM-based training algorithm for RBF networks," *Neural Networks*, vol. 16, no. 1, pp. 69–77, 2003.