

# UNIVERSITÉ PARIS-SUD

ECOLE DOCTORALE DE MATHÉMATIQUES DE LA RÉGION PARIS-SUD  
LABORATOIRE DE MATHÉMATIQUES D'ORSAY

DISCIPLINE : MATHÉMATIQUES (STATISTIQUES)

## THÈSE DE DOCTORAT

Soutenue le 01/04/14 par

**Rémy Fouchereau**

## Modélisation probabiliste des courbes S-N

**Directeur de thèse :** M. Gilles Celeux  
**Co-directeur de thèse :** M. Patrick Pamphile

Directeur de recherche (INRIA)  
Maître de conférences (U-PSUD)

**Composition du jury :**

Président du jury : Pascal Massart  
Rapporteurs : Michel Broniatowski  
André Pineau  
Examineurs : Catherine Duveau

Professeur (U-PSUD)  
Professeur (UPMC)  
Professeur émérite (MINES Paristech)  
Ingénieure (SNECMA)



## Remerciements

Un grand merci tout d'abord à Gilles Celeux pour avoir accepté de diriger mes recherches pendant 3 ans, pour les nombreux conseils donnés et pour les non moins nombreuses discussions dans ton bureau, que ce soit pour parler des modèles de mélange ou de la dernière invasion de coccinelles dans le bâtiment. Je remercie aussi Patrick Pamphile pour sa sympathie, sa disponibilité, ses idées et ses pistes de réflexion, mais aussi pour avoir été à l'instigation de cette thèse et d'avoir cru en moi. Je remercie enfin Catherine Duveau pour m'avoir accueilli à Snecma et avoir pris le temps de suivre régulièrement mon travail ainsi que pour les nombreux conseils avisés, que ce soit en matériau ou en statistiques. J'ai eu de la chance de pouvoir travailler avec vous pendant ces trois années, en tout cas certainement bien plus que vous n'en avez eu avec une tête de pioche telle que la mienne en tant qu'apprenti.

Je remercie André Pineau, Michel Broniatowski et Pascal Massart qui me font l'honneur de participer au jury de cette thèse. Je remercie particulièrement André Pineau pour ses nombreuses remarques et conseils et Michel Broniatowski pour avoir rapporté mon travail.

Je tiens à remercier mes collègues, qu'ils viennent de Snecma ou bien de Paris-Sud et plus particulièrement du 440. J'ai une pensée amicale pour, dans le désordre le plus complet qui me caractérise : Vincent, Patrick, Mélina, Yann, Arnaud, Charles-Antoine, Caroline, Laura, Célia, Elodie, Lucie, Jonathan, Pierre, Shuai, Maud, Christine, Erwan, Yves, Mohammed, Antoine, Raphaël ... [Veuillez insérer ici toute personne oubliée]

Pour conclure, j'adresse un grand merci à ma famille pour leur soutien. Tout particulièrement à Anne-Hélène qui réussit à me supporter au jour le jour, ce qui est quand même assez incroyable il faut bien l'admettre. Et enfin à Camille, qui doit être le seul au monde à pouvoir me faire oublier tous mes soucis en quelques secondes, et ce alors qu'il ne parle pas encore!!!

---



## Résumé

**Résumé** La courbe  $S - N$  est le moyen le plus courant d'analyse et de prédiction de la durée de vie d'un matériau, d'un composant ou d'une structure dans le cadre de la fatigue des matériaux. Cependant, les modèles standards, qu'ils soient basés sur la théorie de la rupture ou sur des modèles probabilistes, n'ajustent pas la courbe dans la totalité sans information sur la microstructure du matériau. Or, cette information provient d'analyses fractographiques souvent coûteuses et rarement disponibles dans le cadre d'une production industrielle. D'un autre côté, les modèles statistiques ne proposent pas d'interprétation physique. Les résultats d'un test de fatigue sont par ailleurs très dispersés, plus particulièrement pour les fortes durées de vie, lieu d'apparition d'un phénomène de bi-modalité. Ces constats sont la raison de la proposition d'un nouveau modèle probabiliste. Celui-ci est composé d'un modèle de mélange spécifique, prenant en compte l'approche apportée par la mécanique de la rupture, sans nécessiter d'information supplémentaire sur la microstructure du matériau. Il utilise le fait que la fatigue peut être vue comme la somme d'un amorçage de fissure et de sa propagation. Les paramètres du modèle sont estimés à l'aide d'un algorithme EM, où la phase de maximisation combine une méthode d'optimisation de Newton-Raphson et une intégration de type Monte-Carlo. Le modèle "amorçage-propagation" offre une représentation parcimonieuse des courbes  $S - N$  dont les paramètres peuvent être facilement interprétés par des ingénieurs matériau. Ce modèle a été testé à l'aide de simulations et appliqué à des données réelles (données sur l'Inconel 718). Ceci nous a permis de mettre en évidence le bon ajustement du modèle à nos données, et ce, pour toutes les déformations disponibles.

**abstract**  $S - N$  curve is the main tool to analyze and predict fatigue lifetime of a material, component or structure. But, standard models based on mechanic of rupture theory or standard probabilistic models for analyzing  $S - N$  curves could not fit  $S - N$  curve on the whole range of cycles without microstructure information. This information is obtained from costly fractography investigation rarely available in the framework of industrial pro-

---

duction. On the other hand, statistical models for fatigue lifetime do not need microstructure information but they could not be used to service life predictions because they have no material interpretation. Moreover, fatigue test results are widely scattered, especially for High Cycle Fatigue region where split  $S - N$  curves appear. This is the motivation to propose a new probabilistic model. This model is a specific mixture model based on a fracture mechanic approach, and does not require microstructure information. It makes use of the fact that the fatigue lifetime can be regarded as the sum of the crack initiation and propagation lifes. The model parameters are estimated with an EM algorithm for which the maximisation step combines Newton-Raphson optimisation method and Monte Carlo integrations. The resulting model provides a parsimonious representation of  $S - N$  curves with parameters easily interpreted by mechanic or material engineers. This model has been applied to simulated and real fatigue test data sets. These numerical experiments highlight its ability to produce a good fit of the  $S - N$  curves on the whole range of cycles.

---

---

# Table des matières

<b>1</b>	<b>Cadre de l'étude</b>	<b>15</b>
1.1	Inconel 718 . . . . .	15
1.1.1	Microstructure de l'Inconel . . . . .	16
1.2	Résistance à la rupture . . . . .	17
1.2.1	Essais de traction . . . . .	18
1.2.2	Essais de fatigue . . . . .	20
1.2.3	Courbe de Wöhler . . . . .	21
1.3	Présentation des données Snecma . . . . .	22
1.4	Conclusion . . . . .	24
<b>2</b>	<b>État de l'art de la modélisation</b>	<b>25</b>
2.1	Modèles de mécanique des matériaux . . . . .	25
2.2	Modèles statistiques . . . . .	28
2.3	Modèles mixtes . . . . .	30
2.3.1	Modèles de mélange et modèle à risques concurrents . . . . .	30
2.4	Exploitation des données Snecma . . . . .	33
2.4.1	<i>Quantile-Quantile plot</i> . . . . .	34
2.4.2	Application à nos données . . . . .	34
2.4.3	Mélange ou risques concurrents . . . . .	35
2.5	Conclusion . . . . .	39
<b>3</b>	<b>Endommagement de l'éprouvette</b>	<b>41</b>
3.1	Mécanisme d'amorçage . . . . .	41
3.1.1	Amorçage sur défauts . . . . .	42
3.1.2	Amorçage en phase I (sur micro-plasticité) . . . . .	43
3.2	Propagation de la fissure et rupture . . . . .	44
3.3	Modèle à deux phases d'amorçage . . . . .	44
3.4	Traitement mathématique de la somme . . . . .	45
3.4.1	Distribution lognormale . . . . .	45
3.4.2	Distribution de somme de lognormales . . . . .	47
3.4.3	Approximation par une autre distribution . . . . .	48

---

---

3.4.4	Approximations numériques . . . . .	49
3.4.5	Approximation des queues de distribution . . . . .	52
3.5	Conclusion . . . . .	53
<b>4</b>	<b>Un modèle de mélange Amorçage-Propagation</b>	<b>55</b>
4.1	Proportion constante . . . . .	56
4.1.1	Estimation des paramètres du modèle . . . . .	56
4.1.2	Résultats sur données réelles . . . . .	61
4.2	Proportion variable . . . . .	62
4.2.1	Proportion affine par morceaux . . . . .	62
4.2.2	Proportion logistique . . . . .	66
4.3	Calibrage de l'algorithme . . . . .	67
4.3.1	Initialisation . . . . .	68
4.3.2	Nombre de simulations Monte-Carlo . . . . .	70
4.3.3	Nombre d'itérations de l'algorithme S-EM . . . . .	70
4.3.4	Nombre de points disponibles . . . . .	71
4.4	Réduction du risque, quel(s) quantile(s) utiliser ? . . . . .	72
4.5	Application à différentes températures . . . . .	74
4.6	Conclusion . . . . .	77
4.6.1	Prédiction . . . . .	77
4.6.2	Classification des données . . . . .	77
4.6.3	Avantages et inconvénients de la modélisation . . . . .	80
<b>5</b>	<b>Prise en compte d'informations de production</b>	<b>83</b>
5.1	Forgeage d'une pièce de turboréacteur . . . . .	83
5.2	données disponibles . . . . .	84
5.2.1	Informations de production . . . . .	85
5.3	Modèle Hiérarchique . . . . .	87
5.3.1	Modèle de mélange d'experts . . . . .	87
5.3.2	Modèle et notations . . . . .	87
5.3.3	Schéma . . . . .	89
5.4	Algorithme EM . . . . .	90
5.4.1	Étape E . . . . .	91
5.4.2	Étape M . . . . .	91
5.4.3	Résultats . . . . .	92
5.5	Modèle ascendant . . . . .	94
5.5.1	Modèle . . . . .	95
5.5.2	Algorithme . . . . .	96
5.5.3	Stabilisation de l'algorithme . . . . .	98
5.5.4	Résultats sur données réelles . . . . .	99
5.6	Conclusion . . . . .	105

---



---

<b>6 Conclusion et perspectives</b>	<b>107</b>
6.1 Conclusion . . . . .	107
6.2 Perspectives . . . . .	107
6.2.1 Equilibrage du plan d'expérience . . . . .	108
<b>A Identifiabilité du modèle de mélange Amorçage Propagation</b>	<b>117</b>
<b>B Méthodes d'optimisations</b>	<b>121</b>
B.1 Méthode de Newton-Raphson . . . . .	121
B.2 Méthode 'BFGS' . . . . .	122
<b>C Résultats supplémentaires de simulation</b>	<b>125</b>

---



---

# Introduction générale

*Remarque préliminaire : Les données présentes dans ce rapport ont été anonymisées afin de conserver leur confidentialité. C'est pourquoi certains graphiques ne posséderont pas de valeurs numériques.*

Le travail présenté dans ce mémoire est issu d'une collaboration entre l'Université Paris-Sud XI et l'entreprise Safran-Snecma.

Safran, conçoit, développe, produit et commercialise des moteurs pour avions civils et militaires, ainsi que pour lanceurs spatiaux. La maîtrise des risques techniques qui peuvent survenir au cours du fonctionnement d'un moteur est donc une préoccupation majeure, que ce soit du point de vue client (avionneurs, compagnies, ...) mais aussi des exigences des autorités de navigabilité (DGAC, DGA, ...). Safran doit garantir le bon fonctionnement de ses moteurs dans le temps et les conditions contractuelles ou réglementaires données. Des stratégies de maintenance sont mises en place pour permettre l'exploitation de matériels et préserver la sécurité des personnes. L'entretien, la réparation et le remplacement du moteur doit donc être conduit sous une double contrainte de performance et de moindre coût. Nécessairement, cette maintenance optimale est basée sur la prévision de la durée de vie des composants du moteur. Ces prévisions se fondent généralement en premier lieu sur une analyse des données du retour d'expérience mais d'aussi de nombreux essais réalisés avant la commercialisation du moteur. Il s'agit là du cadre général pour la fiabilité d'un moteur. Dans le cadre particulier du matériau constituant le moteur, nous n'avons pas facilement accès à ce retour d'expérience.

Les pièces étudiées au cours de cette thèse sont généralement critiques et aucune défaillance (c'est-à-dire fissuration totale) n'est admise en vol. Les données issues d'un accident réel sont très rares (voir Figure 1). Les essais sont donc une des méthodologies usuelles pour obtenir une durée de vie des pièces constituant le moteur. Il s'agit de prélever un petit échantillon du brut de forge<sup>1</sup>, et d'y faire subir des contraintes jusqu'à cassure complète de l'échantillon. Ces tests rentrent dans la famille des tests accélérés de durée

---

1. pièce forgée non usinée

---



FIGURE 1 – Accident Pensacola, Floride, USA, 06 juillet 1996. Eclatement d'un disque de MD88 Delta Air Lines, Moteur P&W JT8D.

de vie. Safran dispose donc d'une base de données conséquente issue d'essais réguliers en production.

Les études de fiabilité des matériaux rencontrent deux difficultés majeures :

- la complexité du processus physique lors de la fissuration et de la propagation de la fissure jusqu'à la rupture de la pièce,
- le nombre important de paramètres qui influe sur la durée de vie totale.

D'autant que de nombreux paramètres influent sur la qualité et la robustesse du matériau. La description du procédé de fabrication d'un moteur est donnée avec la Figure 2. La multiplication des étapes entraîne une multitude de facteurs dont on doit contrôler la variabilité. Le dernier facteur rentrant en jeu dans la durée de vie d'une pièce est les conditions d'utilisation du moteur en vol.

Tous ces facteurs rendent difficile une modélisation complète de la durée de vie.

Ce travail de thèse est avant tout un travail de modélisation, s'inscrivant à la suite de nombreux travaux déjà existants, que ce soit des modèles statistiques ou issus de la mécanique du matériau. Le but principal est l'estimation satisfaisante de grandeurs typiques de fiabilité industrielle, comme des quantiles à 0,1% de défaillance. Par satisfaisante, nous entendons à la fois une estimation fiable, mais aussi une compréhension aisée de la modélisation par les ingénieurs, pouvant permettre l'amélioration du processus de fabrication. Les principales contributions de ce travail sont les suivantes :

- une analyse des données utilisées dans le cadre du suivi ordinaire du matériau par Safran-Snecma. L'augmentation, ces dernières années, du nombre de données au sein de la base, a permis de voir apparaître un
-

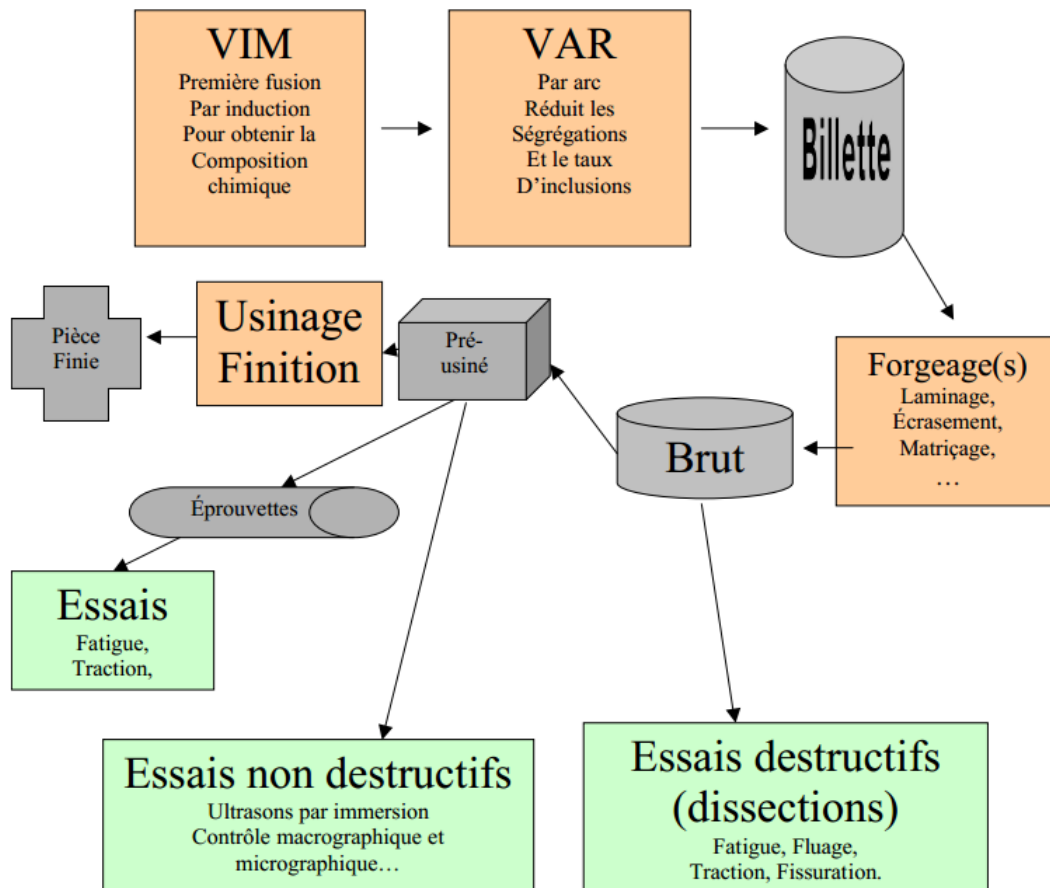


FIGURE 2 – Schéma de production d'une pièce en Inconel 718.

phénomène de bi-modalité qui n'était pas pris en compte jusqu'alors dans les modèles usuels. Par ailleurs, il a fallu relier cette observation, aux connaissances issues de la mécanique des matériaux. Cette partie fait l'objet des trois premiers chapitres de cette thèse.

- Nous avons proposé et mis en place un modèle de mélange « amorçage-propagation », prenant appui à la fois sur l'expertise statistique menée, mais ayant un sens du point de vue de la propagation de fissure au sein du matériau.

Cet objectif parcourt la thèse, appréhender la variabilité observée sur les données d'essais, d'un point de vue mécanique et statistique.

L'apport au niveau du matériau est le traitement de données de fatigue sans connaissance a priori des raisons de rupture, données généralement exploitées par les experts lorsqu'ils s'intéressent aux essais de fatigue.

En effet de nombreux modèles reposent sur la différenciation des causes initiales de fissuration des éprouvettes obtenues après des analyses fractographiques coûteuses. Notre modèle est efficace même sans ces données.

L'apport statistique principal de cette thèse provient ici du fait que le modèle de mélange amorçage-propagation ne constitue pas dans un mélange classique de type :

$$\pi f_1 + (1 - \pi) f_2,$$

où  $f_1$  et  $f_2$  sont des distributions connues.

En effet le modèle que nous proposons traduit la durée de vie  $N$  comme la somme d'une phase d'amorçage  $N_i$  et une phase de propagation  $N_p$  :

$$N = N_i + N_p.$$

Toutefois plusieurs causes d'amorçage peuvent se produire au sein de l'éprouvette. Ainsi la durée de vie sera en fait exprimée de la manière suivante :

$$N = \begin{cases} N_p & \text{si amorçage au premier cycle;} \\ N_i + N_p & \text{sinon.} \end{cases}$$

Puisque la durée de vie en amorçage pour certaines ruptures est négligeable. ( $N_i \simeq 0$ ). Par conséquent notre modèle complet s'écrit :

$$N = \pi(S) f_{N_i} + (1 - \pi(S)) f_{N_i + N_p},$$

Où  $\pi$  est la probabilité d'amorcer au premier cycle.

Ainsi les algorithmes usuels de type Expectation-Maximisation (EM) ont du être adaptés afin de prendre en compte deux particularités :

1. l'utilisation d'une somme de lognormale au sein d'un modèle de mélange ;
2. une proportion  $\pi$  variant suivant la variable déjà utilisée en régression dans les deux modes du mélange.

Cette partie a mené à un travail théorique, puisqu'il a fallu démontrer les bonnes propriétés du modèle (identifiabilité, trouver un équivalent des queues de distribution).

Par ailleurs, le modèle a été testé à la fois sur données simulées et sur données réelles ce qui a permis d'en donner les conditions d'utilisation optimales, cette partie est traitée dans le chapitre 4 ;

- l'extension de la modélisation dans le cadre du suivi de production, ou plus modestement, dans le cadre de la recherche de paramètres influents sur la durée de vie. Ceci fait l'objet du cinquième chapitre.

# Chapitre 1

## Cadre de l'étude

Nous nous intéressons à la durée de vie d'une pièce chaude d'un moteur d'avion. Il s'agit de comprendre les mécanismes et modes de rupture en vol d'une telle pièce. Nous verrons plus particulièrement dans cette partie quels sont les tests mis en œuvre pour étudier les limites mécaniques du matériau, qui sera, dans notre cas, l'Inconel 718.

### 1.1 Inconel 718

L'Inconel 718 a été mis au point en 1959 par Herbert Eiselstein, c'est l'alliage de base Nickel le plus utilisé pour la fabrication de disques de turbines aéronautiques. Un disque doit en effet pouvoir résister à des températures allant de  $450^{\circ}\text{C}$  à  $650^{\circ}\text{C}$  et à des contraintes en fonctionnement pouvant aller jusqu'à 800 MPa dans les zones les plus critiques.

La figure 1.1 montre la résistance à l'oxydation<sup>1</sup> et au fluage<sup>2</sup> de différents matériaux. On constate que seuls les alliages à base principale de Cobalt ou de Nickel peuvent supporter de telles températures. L'alliage 718 a pour avantage, outre ses bonnes propriétés mécaniques<sup>3</sup>, d'être peu cher comparé à d'autres alliages Nickel, notamment le WASPALOY qui fût utilisé jusque dans les années 1980<sup>4</sup>.

---

1. Réaction chimique au cours de laquelle se produit un échange d'électrons, responsable de l'altération du matériau

2. Déformation irréversible non-instantanée d'un matériau soumis à une contrainte constante

3. Propriété caractéristique d'un matériau qui décrit son comportement lorsqu'il est soumis à une ou plusieurs contraintes mécaniques

4. Le changement d'alliage, débuté dans les années 60, s'est achevé après la crise du Cobalt de 1978. L'URSS achetait à cette période tout le Cobalt Zaïrois représentant 60% de la production mondiale.

---

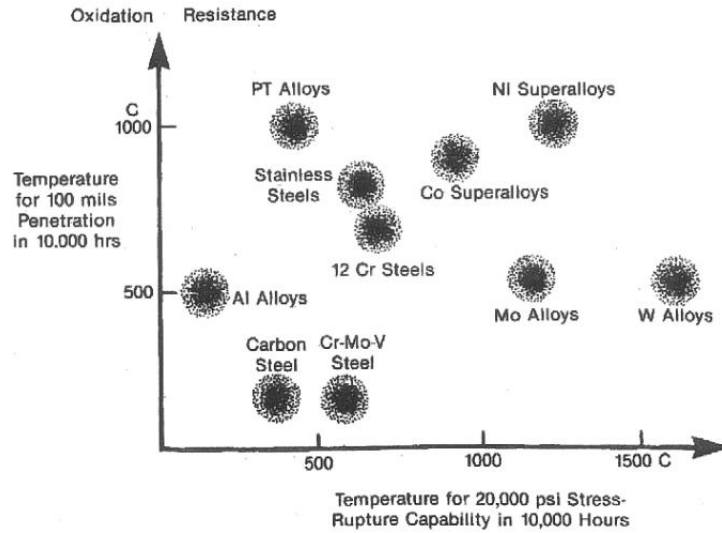


FIGURE 1.1 – Performance en fluage-oxydation des différentes familles de matériaux métalliques Sims et al. [42]

En effet la composition donnée par le Tableau 1.1 montre une forte présence de fer qui permet d'abaisser notablement le prix de cet alliage.

Elements (%)	Ni	C	Cr	Fe	Nb	Mo	Ti	Al
<i>mini</i>	Base	0.02	17	15	4.75	2.80	0.75	0.3
<i>maxi</i>	Base	0.08	21	21	5.50	3.30	1.15	0.7

TABLE 1.1 – Composition de l'alliage 718 utilisé par Snecma

### 1.1.1 Microstructure de l'Inconel

L'inconel est composé de multiples cristaux au sein d'une structure polycristalline. En effet lors du refroidissement du lingot, on observe tout d'abord la formation de sites qui favorisent le regroupement d'atomes. Puis ensuite, le regroupement d'atomes dans le voisinage de ces sites. Ce phénomène se reproduit pour d'innombrables sites dans toute la pièce et forment ce que l'on appelle un réseau cristallin. Lorsque ces réseaux se rencontrent ils ne sont pas alignés, on observe donc une interface entre les deux réseaux, celle-ci est appelée joint de grain (voir Figure 1.2)

La taille de grain est une donnée très importante puisqu'elle influe sur les propriétés mécaniques de l'Inconel. En effet les joints de grain empêchent



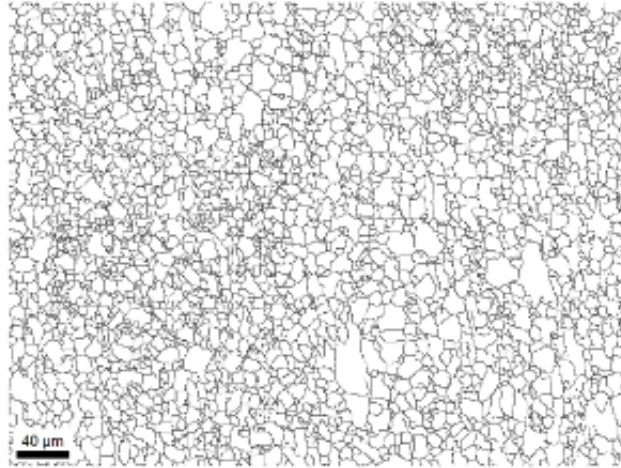


FIGURE 1.2 – Visualisation des joints de grains pour l’inconel 718 Agnoli et al., [2]

les glissements d’un grain à un autre car l’alignement des grains n’est pas identique. Ils sont donc un des facteurs de l’évolution d’une fissure dans le matériau. À composition chimique équivalente, un matériau polycristallin sera plus résistant à la traction (le fait de tirer sur le matériau jusqu’à rupture, voir section suivante) qu’un matériau monocristallin.

Toutefois un matériau polycristallin possède des défauts que l’on retrouve généralement aux joints de matériau. Il s’agit de précipités de type carbures de niobium (NbC) ou nitrures (TiN) créés principalement lors des traitements chimiques appliqués au matériau. Ils en fragilisent la structure et sont souvent les points de départs privilégiés pour une fissure. Bien que la variation des paramètres de forgeage puisse amener à une réduction du nombre et de la taille de ces impuretés, le matériau ne peut être « totalement pur » et de tels défauts se retrouveront dans le produit fini. Cela amène donc une importante problématique de contrôle du procédé de forgeage pour limiter ces défauts, mais aussi à la prise en compte de ceux-ci dans l’optique d’évaluer la résistance du matériau.

## 1.2 Résistance à la rupture

La rupture d’une pièce est tout simplement la séparation en deux ou plusieurs pièces sous l’action de contraintes. Cette rupture peut être partielle, on parlera alors plutôt de brisures, ou complète. On imagine aisément les dégâts que peuvent occasionner une telle séparation pour une pièce aussi critique

qu'un disque de turbine de moteur d'avion, la rupture de cette dernière ne pouvant être contenue dans le carter du moteur. Il est donc indispensable de se doter d'outils afin de déterminer les limites d'utilisation d'un matériau et de prédire sa durée de vie opérationnelle. Cette section présente les deux principaux tests effectués dans ce cadre. Chaque test s'attache à mesurer l'un des deux effets observable sur la Figure (1.3). La première figure étant due à l'application d'une contrainte trop élevée (essais de traction). La deuxième figure représente une rupture du matériau suite au « vieillissement » de celui-ci (essais de fatigue).

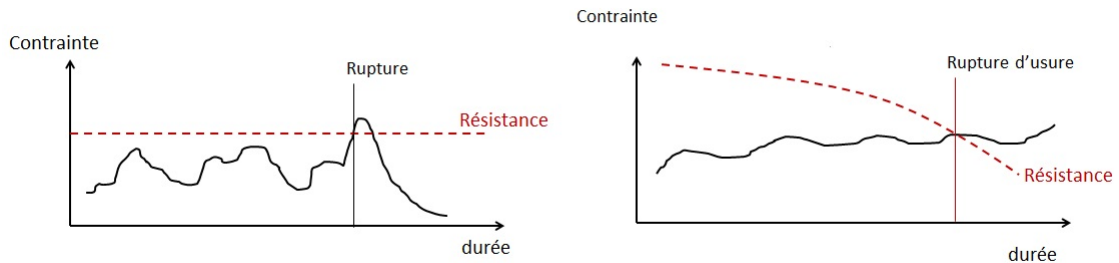


FIGURE 1.3 – Deux types de rupture du matériau

### 1.2.1 Essais de traction

La première démarche nécessaire est de déterminer la contrainte maximale applicable à un matériau avant rupture de ce dernier. C'est là l'objet des tests de traction. Ils consistent à réaliser une traction à température fixée sur une éprouvette de matériau (c'est-à-dire à un échantillon de la pièce avant usinage et finition) jusqu'à la rupture de celle-ci. Pour une pièce ductile<sup>5</sup> on obtiendra généralement le comportement donné en Figure 1.4 :

Ce graphique présente la force  $F$  appliquée à l'éprouvette, en fonction de l'allongement  $\Delta l$  de cette dernière. Dans un premier temps, la déformation est élastique<sup>6</sup>. La courbe s'infléchit ensuite plus ou moins rapidement suivant le matériau, on entre alors dans le domaine de la déformation plastique<sup>7</sup>. Le niveau de cette transition est nommé limite d'élasticité :  $R_e$ . Le maximum est atteint en  $R_m$ , la limite en traction du matériau. Dès lors, on arrive dans

5. La ductilité désigne la capacité d'un matériau à se déformer plastiquement sans se rompre

6. C'est-à-dire réversible. Si on cesse d'appliquer une force à ce moment là, l'éprouvette revient à son état initial

7. Irréversible

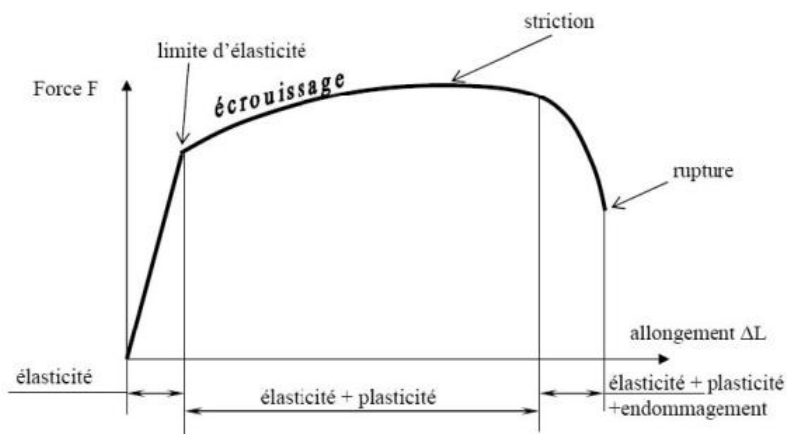


FIGURE 1.4 – Test de traction pour matériau ductile

une phase de striction<sup>8</sup> où la force diminue car la section de l'éprouvette s'amenuise (ce phénomène est visible sur la Figure 1.5). Le test est poursuivi jusqu'à rupture complète de l'éprouvette.

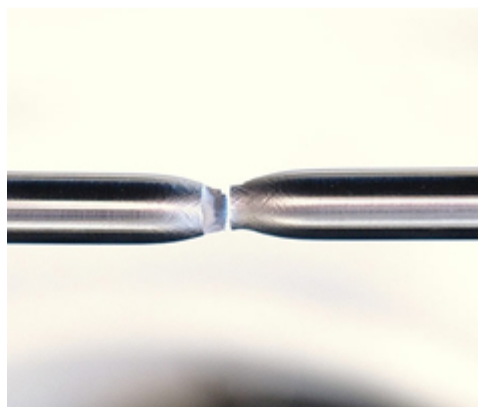


FIGURE 1.5 – Éprouvette rompue suite au test de traction

Il paraît dès lors évident, sauf cas particuliers<sup>9</sup>, que la calibration de la pièce doit être réalisée de telle sorte que la contrainte maximale d'utilisation soit inférieure à  $R_m$ .

8. Réduction locale du diamètre d'une section sous une forte traction

9. cas des dispositifs de sécurité par exemple

### 1.2.2 Essais de fatigue

Le dommage par fatigue concerne les composants soumis à des contraintes variables et qui voient leurs caractéristiques matériaux changer au cours du temps, essentiellement à cause de la formation de fissures. Ce dommage pouvant entraîner la rupture. Ce qui est particulier à la fatigue (et en fait le danger), est que la rupture peut se produire pour des contraintes apparentes relativement faibles (inférieures à  $R_m$ , voire à  $R_e$ ) lorsqu'elles sont répétées un grand nombre de fois.

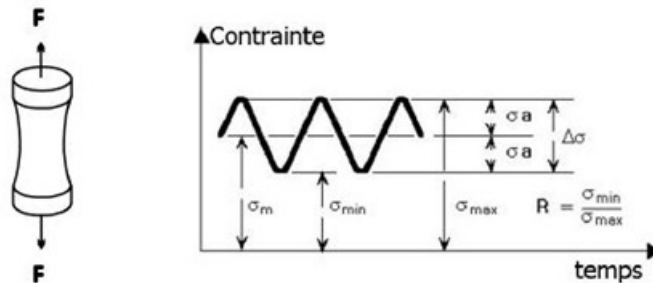


FIGURE 1.6 – Test de fatigue

Afin de tester cet endommagement, les tests de fatigue furent mis en place vers la fin du  $XIX^{eme}$  notamment par August Wöhler. Il s'intéressa au phénomène après la survenue de nombreux accidents de trains par rupture d'essieux, les calculs statiques classiques de résistance des matériaux ne prédisant pas de tels dégâts.

Le *modus operandi* de la réalisation d'un test de fatigue est le suivant : Il s'agit d'appliquer à une éprouvette, une contrainte cyclique à température fixée : l'éprouvette se déforme (allongement/compression) jusqu'à la rupture effective de cette éprouvette. La Figure 1.6 donne les grandeurs usuelles utilisées pour réaliser un tel test :

- $(\sigma_{min}, \sigma_{max})$ , la contrainte minimale (resp. maximale) appliquée à l'éprouvette,
- $\Delta\sigma$  l'amplitude de contrainte pour un cycle,
- $R = \frac{\sigma_{min}}{\sigma_{max}}$ , le rapport de charge du test. Il est à noter que pour cette étude, les rapports de charge pris en compte seront nuls ( $R = 0$ ).

Le comportement d'une éprouvette lors d'un essai de fatigue est bien plus complexe que pour un test de traction (voir les différents faciès de rupture Figure 1.7). Cela vient principalement du fait que le matériau vieillit au

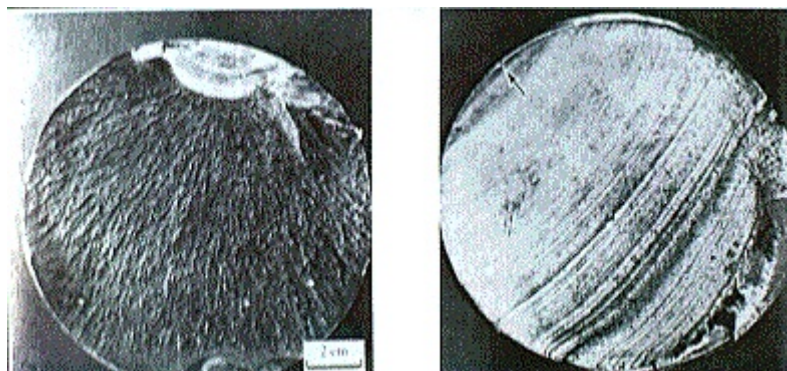


FIGURE 1.7 – Faciès de rupture en fatigue de deux éprouvettes

cours du temps et que les causes de rupture en fatigue sont nombreuses. Par ailleurs, deux possibilités d'asservissement de l'éprouvette sont réalisables :

- la contrainte  $S$  imposée. On contrôle ici, l'effort maximum appliqué à l'éprouvette ;
- la déformation  $\epsilon$  imposée. Où l'on contrôle le coefficient d'allongement de l'éprouvette. S'exprime en pourcentage de la taille de l'éprouvette non chargée.

### 1.2.3 Courbe de Wöhler

Les résultats des tests de fatigue sont regroupés au sein d'une courbe  $S - N$  (Contrainte - Nombre de Cycles), encore appelée courbe de Wöhler, visualisable sur la Figure 1.8.

Les trois zones décrites, sont celles utilisées par les ingénieurs pour décrire trois comportements différents :

- la zone de « fatigue oligocyclique » correspond aux zones de fortes contraintes et de faibles durées de vie. Cette zone est caractérisée par une faible dispersion des données. On observe d'ailleurs ici une limite asymptote à 700 MPa : il s'agit de la limite à rupture du matériau ;  $R_m$  ; à cette contrainte, l'éprouvette casse en un cycle.

*Remarque : Les tests de fatigue sont généralement réalisés bien en dessous de  $R_m$ , ce qui fait que cette borne supérieure n'est pas visualisable sur les courbes  $S-N$  usuelles ;*

- la deuxième zone, « à grand nombre de cycles ». On observe, au fur et à mesure de la baisse de contrainte, une augmentation de la durée de vie ainsi que de la dispersion des données. Il s'agit généralement de la zone d'utilisation de la pièce et cette zone présente donc un grand nombre de données ;

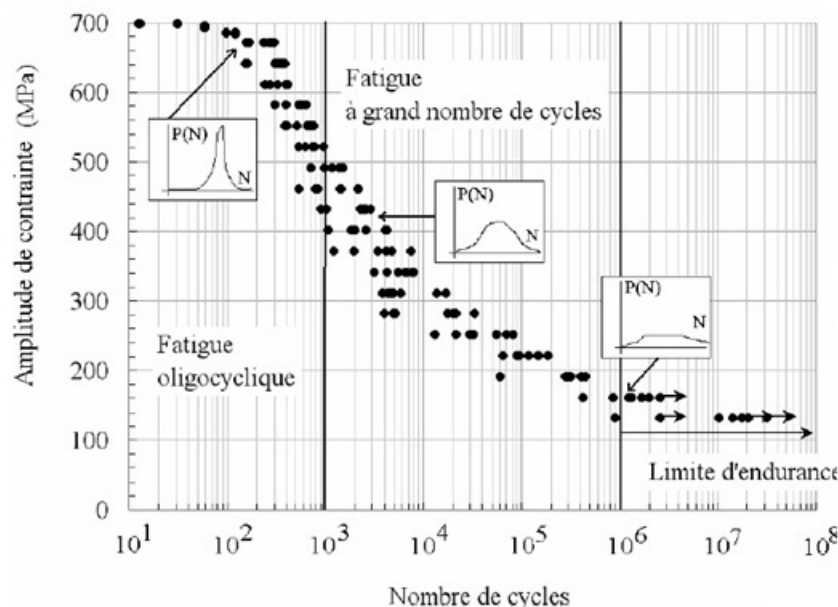


FIGURE 1.8 – Courbe théorique de Wöhler

- enfin, la zone de limite d'endurance, zone où la dispersion des données est très élevée et où certaines éprouvettes peuvent ne pas être rompues, même en  $10^7$  cycles.

En dessous de 100 MPa, les éprouvettes ne rompent plus, la contrainte appliquée étant trop faible. Toutefois l'existence d'une limite d'endurance à  $10^7$  cycles est aujourd'hui remise en cause (par Pittel et al.[35] par exemple), notamment grâce à l'utilisation de tests à haute fréquence qui permettent d'augmenter le nombre de cycles réalisables.

Les principales utilisations de la courbe  $S - N$  sont :

- dimensionner la pièce, et ainsi pouvoir fournir une durée d'utilisation de la pièce. Elle intervient donc dans le calcul du prix de la pièce ;
- contrôler la qualité en construisant un intervalle de tolérance en dessous duquel les pièces sont jugées non conformes.

### 1.3 Présentation des données Snecma

La base de données Snecma comporte environ 4000 points de fatigue pour des tailles de grains allant de 7 à 13 ASTM<sup>10</sup> principalement. Notre étude

<sup>10</sup>. *American society for testing and material*, et, par extension, la norme ASTM E112 mise au point pour les tailles de grain.

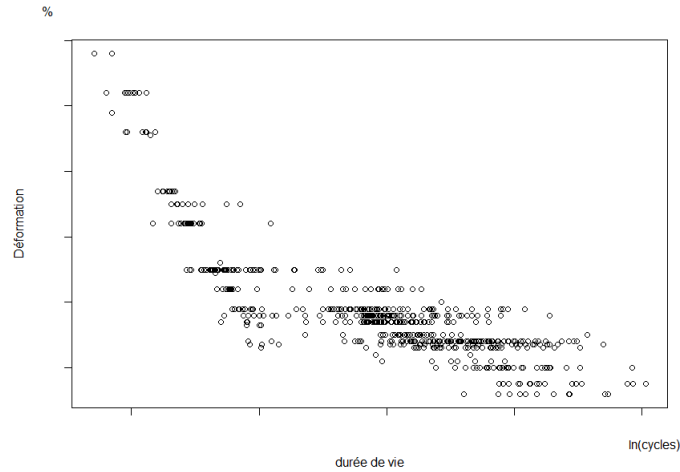


FIGURE 1.9 – 450°C, visualisation de la courbe S-N issue des données snecma.

portera sur **les tailles de grains de 9,5 à 13 ASTM** composée d'environ 2000 données. La répartition de ces points en fonction de leur température d'essai est disponible dans la Table 1.2

température °C	Nombre de points
20	70
150	30
200	500
350	250
450	400
550	600

TABLE 1.2 – Nombre de points de la base en fonction de la température

Les températures testées correspondent au domaine de fonctionnement d'un moteur. Ainsi, 20°C correspond à la température du moteur au sol au moment du démarrage et 450-550°C sont les températures atteintes en fonctionnement. Ces températures élevées sont donc les plus testées puisqu'elles correspondent à la plage de températures en vol.

La courbe de Wöhler obtenue pour une température de 450°C est visualisable sur la Figure (1.9). On constate que la dispersion des points semble plus importante que celle présentée sur la courbe type (voir Figure 1.8). De plus, la répartition des niveaux de déformation testés n'est pas équilibrée : ainsi les faibles niveaux de déformations sont mieux représentés. Par ailleurs on

constate des zones sans niveau de contrainte testé, ce qui contribue à n'avoir qu'une représentation partielle de la courbe de Wöhler.

## 1.4 Conclusion

Le premier objectif de cette thèse sera donc d'obtenir une modélisation cohérente de nos données de fatigue pour l'Inconel 718. Le tout afin de fournir un quantile théorique à 0,1% qui permettra, l'évaluation de la durée de vie maximale en vol de la pièce composant le moteur d'avion. On perçoit déjà dans la forme du nuage de points les futures difficultés rencontrées au cours de cette thèse. Celles-ci sont dues principalement à l'hétérogénéité des données et au faible nombre d'informations disponibles, au vu de la complexité des phénomènes physiques mis en jeu.

De nombreux modèles tant statistiques que mécaniques ont été proposés précédemment. Nous proposons d'en donner un aperçu dans le chapitre suivant.

---



## Chapitre 2

# État de l'art de la modélisation

Cette partie relate des différents modèles utilisés habituellement dans le cadre de la fatigue des matériaux. On retrouve tout d'abord ceux issus de la mécanique de la rupture. Ils s'intéressent à la progression de la fissure, de sa formation jusqu'à la rupture, et cherchent donc à modéliser le plus fidèlement possible chaque phase d'évolution de la fissure. On retrouve ensuite les modèles purement statistiques, exprimant  $\log N$  (le log de la durée de vie) en fonction de  $S$  (la contrainte). Enfin, nous verrons les modèles essayant de mêler les deux approches, et étudierons les forces et les faiblesses de ces modélisations.

*Remarque : Il est à noter que la fatigue est la composante de la déformation élastique et plastique appliquée à la pièce. Cette plasticité est toutefois difficile à mesurer et nous n'en possédons pas d'estimation dans notre base de données. Par ailleurs, à déformation égale tout au long du test, la contrainte appliquée à l'éprouvette diminue au fur et à mesure par effet d'adoucissement du matériau. Là encore, l'évolution de cette contrainte n'est pas disponible. Ainsi nous nous contenterons dans nos modèles des valeurs  $\epsilon$  de la déformation totale et  $N$  du nombre de cycles. Le graphique 2.1 montre la dispersion des modules d'Young au sein de notre base et illustre la difficulté d'obtenir des mesures précises des paramètres du matériau.*

### 2.1 Modèles de mécanique des matériaux

La rupture de fatigue se caractérise par l'apparition de microfissures au sein du matériau, puis, sous l'effet de contraintes cycliques, ces microfissures donnent naissance à une fissure qui peut se propager jusqu'à la rupture.

Les modèles de mécanique des matériaux permettent de décrire précisément les phénomènes de fissure de leur amorçage jusqu'à la rupture.

---

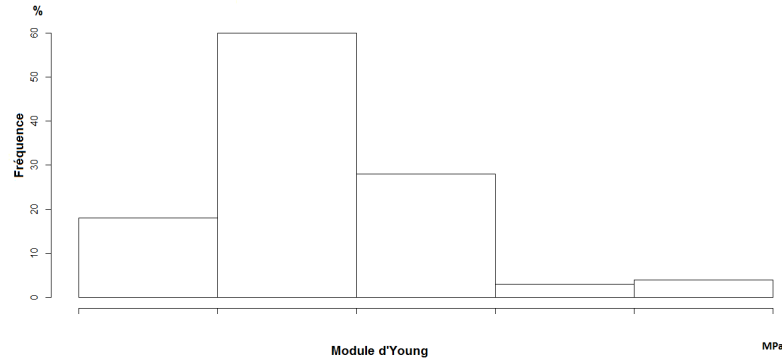


FIGURE 2.1 – Dispersion du module d'Young au sein de notre base à 450°C pour une déformation de 0,36.

La Figure 2.2 montre l'évolution de la vitesse de propagation de la fissure ( $\frac{da}{dN}$ ) en fonction de  $\Delta K$  la variation du facteur d'intensité de contrainte<sup>1</sup> ( $a$  représente donc la longueur de fissure).

On observe trois domaines principaux sur ce graphique :

I Le domaine I correspondant à de la micro-propagation possédant deux types de comportement : « fissures longues » (trait plein) et fissures courtes (trait pointillé). Pour les « fissures longues », on observe une rapide décroissance de la propagation jusqu'à une valeur seuil de  $\Delta K$  correspondant à la limite où la progression de la fissure n'est plus mesurable :  $\Delta K_{seuil}$ . Une « fissure courte » dans cette zone, possédera une vitesse de propagation nettement supérieure à celle observée pour les « fissures longues » à  $\Delta K$  équivalent. Cette partie est difficile à analyser au vu de la faible taille des fissures mises en jeu, mais aussi de l'aspect non régulier de la propagation à ces faibles niveaux de contraintes. La fissure peut en effet s'arrêter, on parlera de fissure courte non propageante. C'est le domaine le plus influencé par la microstructure du matériau. De nombreux modèles ont été proposés pour  $\frac{da}{dN}$ , le modèle de Tomkins [47], ou des adaptations du modèle de Paris. Toutefois il n'existe pas encore de modèles universels qui traiteraient les différentes sortes de fissures courtes d'un seul bloc.

II Ce domaine est le plus simple à modéliser, la courbe présente ici une partie linéaire. La loi de Paris est celle régissant la propagation dans ce domaine (Paris et Erdogan [33]). Soient  $C$  et  $m$  des constantes dépendant

1. En  $\text{MPa}\cdot\text{m}^{1/2}$ . Caractérise l'ensemble du champ des contraintes autour de la pointe de fissure

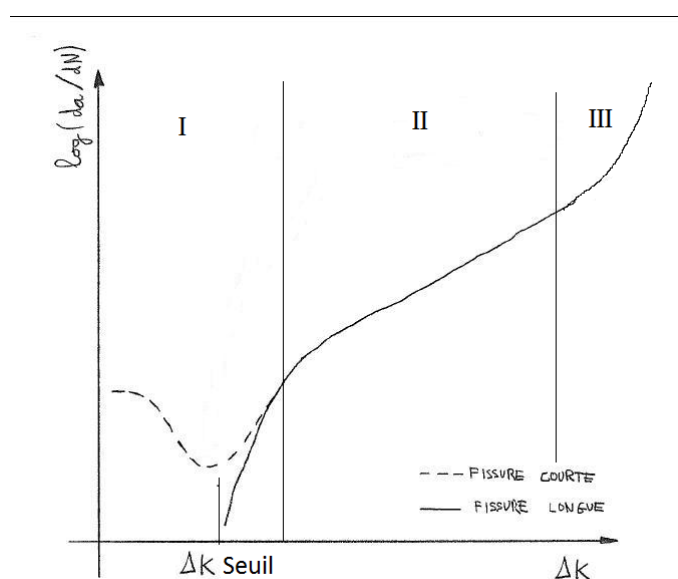


FIGURE 2.2 – Évolution de la vitesse de propagation en fonction de la variation du facteur d'intensité de contrainte sur un cycle.

du matériau,  $N$  le nombre de cycles et  $\Delta K$  la variation du facteur d'intensité de contraintes. La loi de Paris s'écrit alors :

$$\frac{da}{dN} = C(\Delta K)^m. \quad (2.1)$$

III Ensuite, lorsque l'on approche du paramètre de ténacité<sup>2</sup> du matériau, la vitesse s'accélère et entraîne une rupture rapide. Cette dernière zone est la moins étudiée, en effet celle-ci conduit inéluctablement à la rupture rapide et représente une faible proportion de la durée de vie totale de la pièce contrairement aux phases I et II.

Les modèles proposés dans le cadre d'une étude de la fissuration sont complexes, notamment en ce qui concerne la micropropagation. En effet le comportement erratique de la propagation dans cette zone le rend difficilement modélisable, d'autant que deux comportements totalement différents apparaissent, celui des fissures « longues » et celui des fissures « courtes ».

2. Quantité d'énergie qu'un matériau peut absorber avant de casser

## 2.2 Modèles statistiques

*Remarque : afin de garder les notations statistiques usuelles,  $\epsilon$  ne représentera pas dans cette section la déformation mais les fluctuations aléatoires.*

Cette section regroupe les principaux modèles mathématiques utilisés pour décrire les courbes de Wöhler. Les modèles présentés ici sont des modèles du type :

$$\log N = f(S) + \sigma\epsilon,$$

où  $f(s)$  représente une tendance (durée de vie moyenne ou médiane) qui dépend de la contrainte,  $\epsilon$  représente les fluctuations aléatoires non assignables et  $\sigma$  la dispersion des données. Tous ces modèles n'utilisent donc aucune information physique sur les données.

*Remarque : Il est à noter que tous les modèles ont été transformés de manière à ce qu'ils soient tous de forme comparable. Le modèle de Bastenaire par exemple, exprime  $S$  en fonction de  $\log N$  sous sa forme originelle.*

Modèle	Équation fonctionnelle
Wöhler (1870)	$\log N = a + b S$
Basquin(1910)	$\log N = a + b \log(S)$
Strohmeyer (1914)	$\log N = a + b \log(S - S_0)$
Palmgren (1924)	$\log(N + d) = a + b \log(S - S_0)$
Weibull (1949)	$\log(N + d) = a + b \log\left(\frac{S-S_0}{S_t-S_0}\right)$
Bastenaire (1972)	$\log(N + d) = \frac{a \exp^{-c(S-S_0)}}{S - S_0}$
Spindel-Haibach (1981)	$\log N = a + b \log(S) + c \log\left(1 + \left(\frac{S}{S_0}\right)^{-2\alpha}\right)$
Castillo et al.(1985)	$\log N = a + \frac{b}{\log(S) - \log(S_0)}$
Pascual and Meeker (1999)	$\log N = a + b \log(S - S'_0) ; S'_0 \sim \mathcal{N}(\alpha, \beta)$

TABLE 2.1 – Modèles tirés en partie du livre de Castillo et Canteli [12].

Le tableau 2.1 regroupe les principaux modèles statistiques utilisés. Les premiers modèles se contentent de travailler aux faibles contraintes ce qui explique leur forme, la courbe présentant une partie linéaire importante dans cette zone. Au fil du temps, la forme de la courbe s'est améliorée afin d'épouser la forme de la courbe de Wöhler sur un plus grand nombre de niveaux de contraintes. Le dernier modèle proposé étant celui de Pascual et

Meeker [34]. Ce modèle est le seul à prendre en compte une limite d'endurance aléatoire.

*Remarque : En plus de ces modèles, on trouve de nombreux modèles utilisant un découpage de  $S$  afin de réaliser des régressions par morceaux. On pourra citer par exemple l'article de Pollack et al. [36].*

Tous ces modèles ont toutefois un gros défaut : ils ne s'adaptent que partiellement à nos données. En effet on remarque que tous ceux-ci sont composés d'une unique équation de courbe et donc d'une seule distribution. La Figure 2.3 présente le modèle de Pascual-Meeker appliqué aux données Snecma.

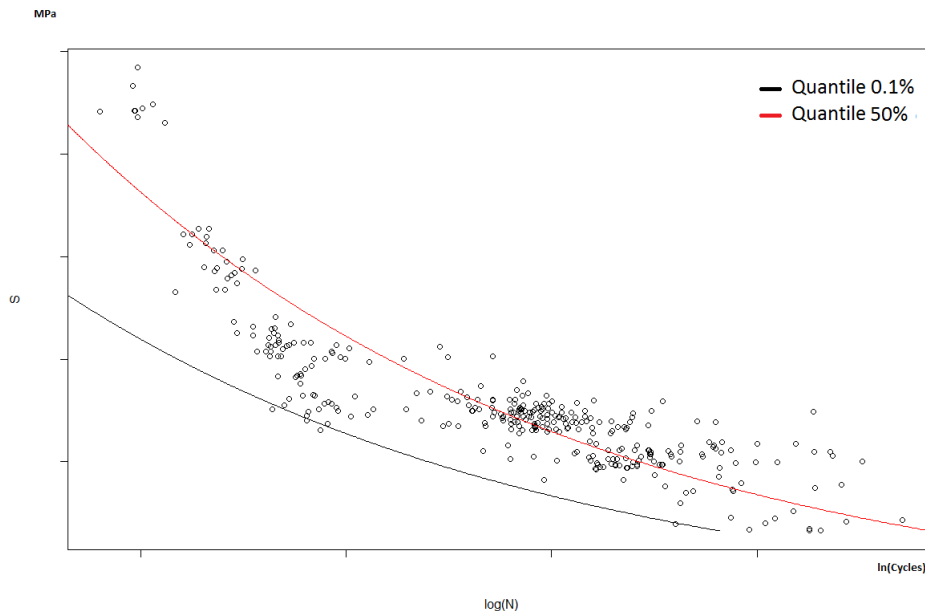


FIGURE 2.3 – Modèle de Pascual-Meeker appliqué aux données Snecma

Outre la mauvaise estimation notamment pour les fortes déformations. On constate un petit groupe de données à 600 MPa qui sort du quantile à 0,1%. Or si une sortie de point peut s'expliquer par le seul fait de la présence d'aléa dans nos données, le fait d'avoir tout un groupe de données qui ressortent du quantile au même endroit tend à montrer un défaut de modélisation dans cette partie-ci de la courbe. Les modèles statistiques classiques ne peuvent donc pas être utilisés tels quels dans notre cas. Bien qu'ils puissent être source d'inspiration pour la suite.

## 2.3 Modèles mixtes

Les modèles présentés ci-dessous sont :

- soit des modèles statistiques nécessitant une information supplémentaire à la seule connaissance de  $S$  et de  $\log N$ .
- soit des modèles de mécanique des matériaux rajoutant une importante partie probabiliste dans le but de traiter la dispersion et donc l'aléa des données en présence d'un grand nombre d'éprouvettes.

Cette section a pour but de présenter des modèles pouvant être mis en place à l'interface de la statistique et de la mécanique des matériaux.

### 2.3.1 Modèles de mélange et modèle à risques concurrents

Les modèles présentés font l'hypothèse d'un comportement différent pour les éprouvettes en fonction de l'amorçage de leur fissure : en surface ou à cœur. Nous ne pouvons pas appliquer directement ces méthodes puisque notre base de données ne présente que très peu d'information de ce type.

#### Modèles de risques concurrents

Ces modèles sont issus des travaux de Bomas [9]. Ici l'auteur fait l'hypothèse qu'il y a compétition entre deux types de rupture, celle associée à un amorçage sur défaut et les autres. Pour chaque catégorie, il met en place un modèle de Basquin (voir Table 2.1) :

$$\log N_K = a_K + b_K \log S + \epsilon_K,$$

et fait entrer les deux distributions en compétition, ce qui lui permet d'obtenir le modèle suivant :

$$\log N = \min(\log N_{\text{defaut}}, \log N_{\overline{\text{defaut}}}),$$

où  $N_{\text{defaut}}$  représente la durée de vie liée à un amorçage sur défaut et  $N_{\overline{\text{defaut}}}$  un amorçage autre.

Chaque partie du modèle est estimée indépendamment puisque l'auteur possède la connaissance du type de défaut. Toutefois, dans l'article, ce modèle n'est pas utilisé pour décrire entièrement la courbe de Wöhler mais uniquement des contraintes allant de 500 à 650 MPa (voir Figure 2.4). Se pose donc la question de la validité du modèle pour des contraintes supérieures ou inférieures à ce domaine d'utilisation. On observe notamment une concavité du modèle utilisé qui ne correspond à aucune réalité pour ce qui concerne les données de notre base. De plus, on notera la très faible présence de points

appartenant à la catégorie des amorçages sur défaut. La précision de l'estimateur de ce mode est donc assez faible dans le cas étudié.

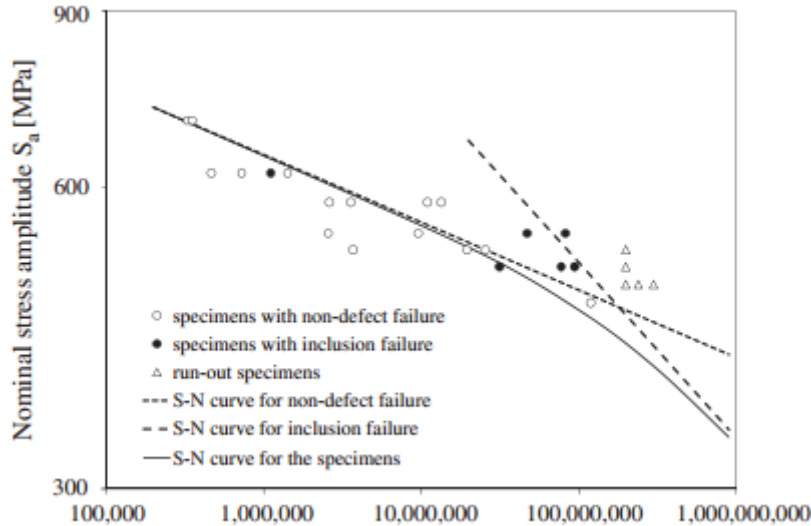


FIGURE 2.4 – Modèle à risques concurrents pour courbe S-N. Bomas [9]

### Modèles de mélange

Plusieurs modèles de mélange ont été proposés dans la littérature. On pourra notamment citer Sakaï et al. [39] qui propose un modèle de mélange de lois de Weibull, ou encore Weixing et Shejieng [48] avec un mélange de lognormales. Ici la distinction entre les deux modes se fera suivant l'analyse d'un amorçage d'une fissure interne ou à cœur. La durée de vie s'exprime donc sous la forme :

$$N = \pi f_{N_{surface}} + (1 - \pi) f_{N_{interne}}.$$

Avec  $\pi$  représentant la probabilité que l'amorçage ait eu lieu en surface et où  $f$  représente une distribution de Weibull (ou lognormale). Ce modèle décrit bien toute la courbe sur une grande plage de contraintes différentes. Toutefois, le calcul des estimateurs pratiqué par Sakaï n'est pas réalisé directement. Il estime d'abord les paramètres des lois de Weibull à chaque niveau de déformation, avant de proposer un modèle plus général en se servant uniquement des paramètres obtenus à chaque niveau. L'auteur ne propose donc pas une estimation directe des paramètres sur les données complètes ce qui donne un estimateur peu consistant au final.

Par ailleurs, Sakai suppose connu le fait que l'on sache, pour chaque éprouvette la nature de l'amorçage. Cette observation nécessite une analyse fractographique de la surface de rupture par microscope à balayage électronique. Cette analyse microscopique est coûteuse et n'est donc pas effectuée de manière systématique. L'estimation de  $\pi$  sur les données Snecma n'est donc pas évidente.

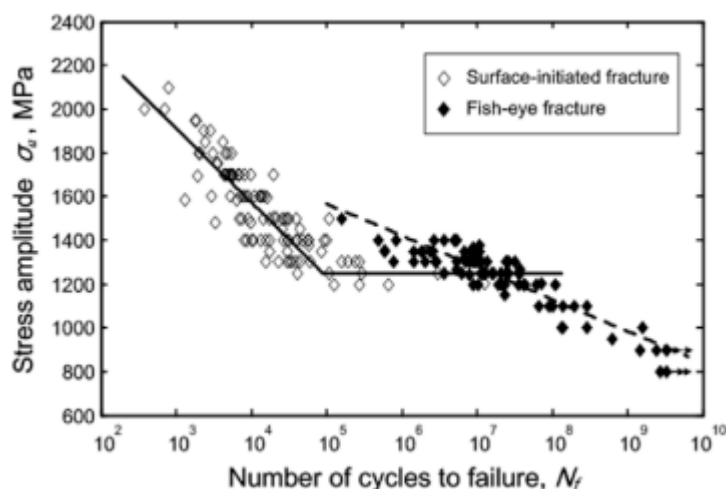


FIGURE 2.5 – Modèle de mélange Sakai. [39]. Acier pour roulements à haute teneur en carbone.

Cependant, les données exploitées par Sakai ressemblent fortement à celles étudiées dans le cadre de cette thèse. Ainsi même si la méthode n'est pas exploitable telle quelle, la modélisation principale proposée au cours du chapitre 4 reprendra l'idée d'un mélange associé à deux modes de rupture différents.

### Modèles probabilistes liés à la mécanique des matériaux.

L'intégration d'une partie probabiliste dans la construction d'un modèle de propagation vient souvent de l'aléa observée sur les tailles de particules. (cf. Alexandre [4], Thieulot-Laure [46]). Ainsi la modélisation probabiliste de l'amorçage tient principalement à une modélisation des défauts (taille et probabilité de présence), mais aussi au domaine d'activation. Ainsi un défaut nocif à 600 MPa ne le sera pas nécessairement à 450 MPa.

Les modèles probabilistes sont donc généralement réalisés en trois étapes :

1. Premièrement, identification des paramètres des distributions de défauts ;
2. Ces défauts initiaux sont ensuite simulés un grand nombre de fois à différentes contraintes ;



3. Les modèles classiques de propagation ou de micropropagation sont ensuite appliqués à chaque point simulé ce qui permet d'obtenir au final des courbes de fatigue.

Toutefois le fait que ce sont les données d'entrée du modèle (taille des défauts) et non pas les données de sortie ( $S$  et  $N$ ) qui sont calibrées, fait que l'on observe un écart entre la durée de vie prédite par cette méthode et la réalité observée. Ainsi, même si les modèles utilisés tendent à se rapprocher de l'expérimentation, il reste des écarts dus à deux phénomènes :

- la complexité des interactions entre les divers phénomènes physiques mis en jeu dans le cadre de la fatigue rend difficile l'adéquation exacte entre une simulation qui essaye de retracer le plus fidèlement possible l'évolution d'une fissure et la réalité de cette évolution ;
- l'identification des modèles sur un groupe restreint de données. En effet chaque partie des modèles est évaluée indépendamment des autres. Ainsi on évaluera dans un premier temps les paramètres liés à la taille des défauts. Une autre étude donnera une courbe de propriété du matériau. Enfin une dernière étude donnera la corrélation entre la taille de grain et la propagation. Le problème étant qu'à chaque identification est associée une erreur incompressible sur les estimateurs (sauf à posséder plus de points). Chaque erreur d'estimation est donc cumulée lors de l'obtention du modèle final.

## 2.4 Exploitation des données Snecma

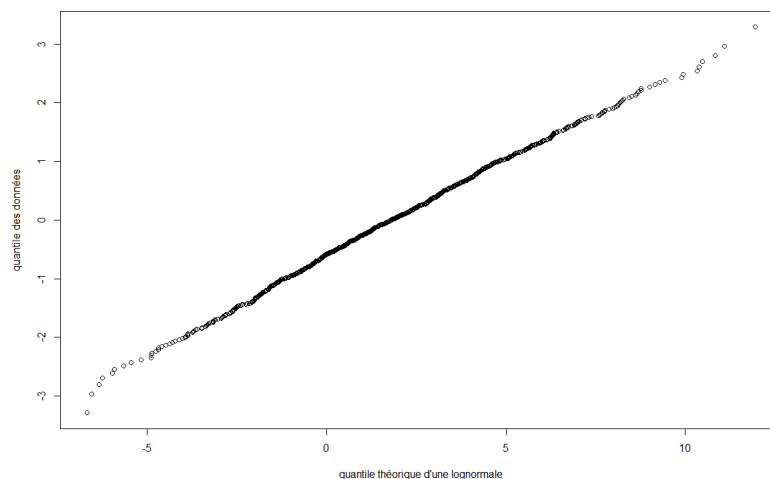
Le point de vue adapté au cours de cette thèse sera de trouver la modélisation la plus fidèle possible aux données de sortie, c'est-à-dire les durées de vie ( $N$ ) observées. Ainsi, la première phase d'étude portera sur l'observation des données issues de la base Snecma, étude qui conditionnera le reste de la modélisation proposée.

Les essais Snecma sont réalisés à déformation  $\epsilon$  fixée, puis au bout de 80000 cycles, l'essai passe en contrainte  $S$  imposée, techniquement plus rapide à réaliser. Les modèles statistiques classiques s'écrivent :

$$\log N = f(\epsilon) + \alpha,$$

$N$  représentant la durée de vie,  $\alpha$  l'aléa. Nous cherchons dans un premier temps à identifier cet aléa. On utilise pour cela une méthode graphique, le *quantile plot*, décrite ci-dessous.

---

FIGURE 2.6 – Visualisation du *quantile plot* d'une loi logNormale

### 2.4.1 *Quantile-Quantile plot*

Le *quantile plot*, est une méthode de référence qui permet d'effectuer l'adéquation visuelle entre un jeu de données et une distribution (normale, log-normale, Weibull ...).

La représentation de  $t$  contre  $F^{-1}(t)$  peut être linéarisée en trouvant une transformation adéquate. Dans le cas lognormal, la fonction à appliquer est :

$$\Phi^{-1}(p) = \frac{\log(t_p) - \mu}{\sigma}, \quad (2.2)$$

où  $\Phi^{-1}(p)$  représente le  $p$ -ième quantile d'une distribution normale centrée réduite. Ce qui implique que  $\log(t_p)$  contre  $\Phi^{-1}(p)$  est représenté par une ligne droite.  $\frac{-\mu}{\sigma}$  donne l'ordonnée à l'origine de la droite tandis que la pente est donnée par  $\frac{1}{\sigma}$ . Un exemple de quantile plot est donné par le Figure 2.6 pour une loi lognormale.

### 2.4.2 Application à nos données

Chaque symbole du graphique 2.7 représente des essais pour un niveau de déformation différent.

1. Pour les forts niveaux de déformation (faibles durées de vie), on observe un bon ajustement des points par une droite. Une seule lognormale semble convenir pour la modélisation.

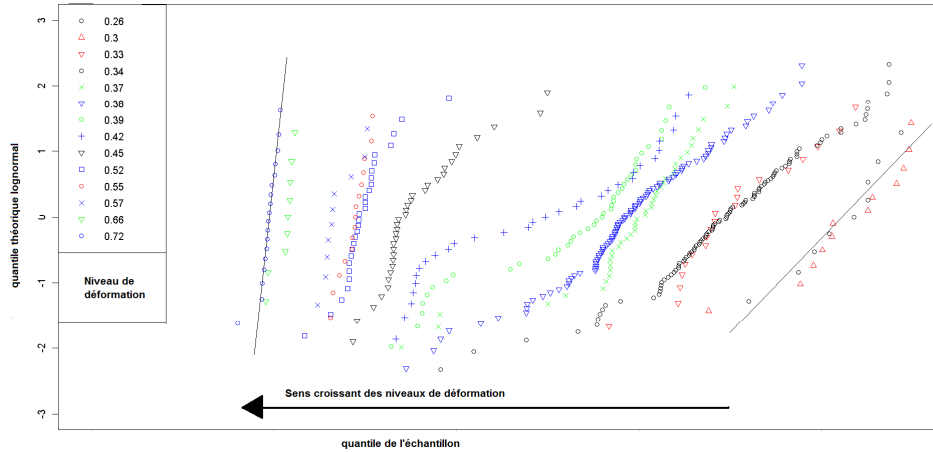


FIGURE 2.7 – quantile plot de la distribution empirique de  $\log N$  contre une distribution lognormale en fonction de la déformation pour une température de  $450^{\circ}\text{C}$ .

2. La même remarque peut être réalisée pour les plus faibles niveaux de déformation (fortes durées de vie), bien que l'ajustement linéaire soit moins bon.
3. Pour ce qui est des niveaux intermédiaires, une seule distribution ne suffit plus. On observe un première partie linéaire qui a la même pente que celle des forts niveaux de déformation (1), puis un plateau, et enfin une autre partie linéaire qui est parallèle à ce qui se passe aux faibles niveaux (2).

Ce phénomène nous indique qu'il y a donc deux distributions des données : une liée au fort niveau de déformation et l'autre aux faibles. Nous pouvons donc proposer soit un modèle de mélange, soit un modèle à risques compétitifs, modèles usuels lorsque l'on observe comme ici plusieurs groupes.

### 2.4.3 Mélange ou risques concurrents

Cette partie a pour but de déterminer la meilleure modélisation à appliquer sur nos données en fonction de considérations sur les *quantile plot* observés.

#### Risques concurrents :

Les modèles à risques compétitifs ont une signification physique particulière. Il s'agit d'une modélisation dite du « maillon faible ». Cette

modélisation, comme son nom l'indique, représente la structure comme une chaîne formée de maillons. Prévoir la rupture de la structure revient donc à prévoir la rupture du maillon le plus faible ou le plus sollicité. La structure survit si aucun des éléments la composant ne rompt. En supposant l'indépendance des durées de vie de chaque maillon  $T_1, \dots, T_n$  on peut écrire que :

$$P(T > t) = P(\min(T_i) > t),$$

où  $T$  représente la durée de vie de la structure complète. Il s'agit donc d'une loi du minimum. Au vu du nombre de micro-fissures susceptibles d'apparaître au sein d'une éprouvette, cela semble être la loi naturelle de nos données en première approche.

Toutefois, les modèles à risques concurrents ne sont pas adaptés à la forme de nos données. Pour cela nous allons démontrer qu'un modèle à risques concurrents donne un « quantile plot » convexe, ce qui n'est pas le cas de nos données. Le cas lognormal qui nous intéresse n'est pas démontré mais des simulations montrent que le comportement est identique au cadre des distributions de Weibull. La première étape est donc de calculer l'équation d'un risque concurrent pour un *quantile plot*.

– Équation :

Le quantile plot pour une distribution de Weibull est le suivant :

$$\log[-\log(1-p)] = \beta(\log(t_p) - \log(\eta)).$$

Il s'agit donc de tracer  $\log(t_p)$  contre  $\log[-\log(1-p)]$ . Avec  $-\log(1-p)$  représentant le  $p$ -ième quantile d'une distribution de Weibull(1,1).

Nous souhaitons maintenant savoir quelle équation de courbe nous obtiendrons dans le cadre d'un modèle à risques concurrents de lois de Weibull. La probabilité  $p$  sera donc remplacée par la probabilité du modèle compétitif à deux distributions de Weibull, paramétrées par respectivement  $(\eta_A, \beta_A)$  et  $(\eta_B, \beta_B)$ . La fonction de répartition d'un modèle à risques concurrents sera notée  $F$ . On a

$$p = F(t) = P(T < t) = 1 - P((T_A < t) \cap (T_B < t)).$$

Les risques sont considérés indépendants :

$$1 - p = W_A(t)W_B(t),$$

avec  $W_A(t) = 1 - e^{-(\frac{t}{\eta_A})^{\beta_A}}$ .

Après réécriture on a :

$$p = 1 - e^{-(\frac{t}{\eta_A})^{\beta_A}} e^{-(\frac{t}{\eta_B})^{\beta_B}}.$$

Pour un modèle à risques concurrents, on peut écrire :

$$g(t) = \log[-\log(e^{-\frac{t}{\eta_A} \beta_A} e^{-\frac{t}{\eta_B} \beta_B})],$$

que l'on peut réécrire, après simplification :

$$g(t) = \log\left[\left(\frac{t}{\eta_A}\right)^{\beta_A} + \left(\frac{t}{\eta_B}\right)^{\beta_B}\right]. \quad (2.3)$$

*Remarque : nous avons exprimé ici  $\log[-\log(1-p)]$  en fonction de  $t$ . Or le Weibull-plot trace cette courbe en fonction de  $x = \log(t)$ . Nous posons donc pour la suite  $t = e^x$ , ce qui donne  $(\frac{t}{\eta_A})^{\beta_A} = e^{\beta_A(x - \log[\eta_A])}$ .*

L'équation de la courbe d'un quantile plot de risques concurrents est donc la suivante :

$$f(x) = \log[e^{\beta_A(x - \log[\eta_A])} + e^{\beta_B(x - \log[\eta_B])}] \quad (2.4)$$

- Convexité de la courbe des risques concurrents : Pour démontrer la convexité de la fonction (2.4) nous allons simplement montrer que sa dérivée seconde est positive sur  $\mathbb{R}_+^*$ .  
On posera  $z_A = e^{\beta_A(x - \log[\eta_A])}$ ,

$$f'(x) = \frac{\beta_A z_A + \beta_B z_B}{z_A + z_B},$$

$$f''(x) = \frac{(\beta_A^2 z_A + \beta_B^2 z_B)(z_A + z_B) - (\beta_A z_A + \beta_B z_B)^2}{(z_A + z_B)^2}.$$

Ce qui, après simplification, donne :

$$f''(t) = \frac{z_A z_B ((\beta_A - \beta_B)^2)}{(z_A + z_B)^2}. \quad (2.5)$$

Or  $z_A, z_B$  sont positifs, la dérivée seconde est bien positive et donc un modèle à risques concurrents présenté sur un « quantile plot » aura une forme convexe. Nos données ne semblent donc pas pouvoir être modélisées par un tel modèle puisque le palier observé pour les données à moyennes déformations contredit la convexité.

On visualise sur le graphique 2.8 la courbe représentant un modèle à deux risques concurrents lognormaux sur un *quantile plot* de lognormale. On observe donc bien le phénomène de convexité démontré dans le cadre de deux lois de Weibull.

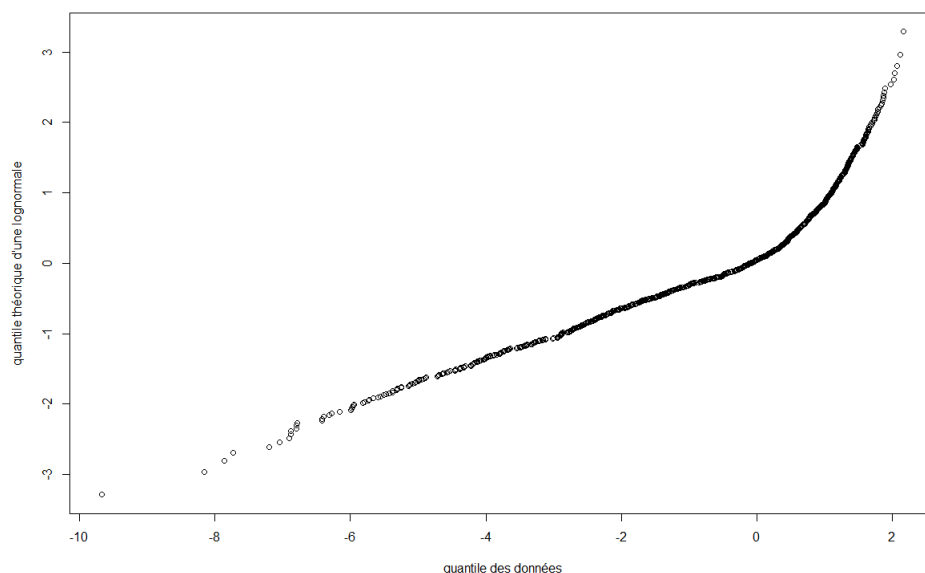


FIGURE 2.8 – Quantile plot d'un modèle de risques concurrents de logNormales

### Modèle de mélange :

L'étude du « quantile plot » d'un modèle de mélange est plus complexe que celle des risques concurrents. Toutefois, nous observons sur des données simulées que l'on peut retrouver une courbe de type escalier à l'aide d'une distribution de ce type.

*Remarque : Les deux droites représentent le quantile plot que nous aurions obtenus si nous n'avions eu que des points suivant une lognormale A ou que des points suivant une lognormale B. La courbe représente le mélange de ces deux composantes (voir Figure 2.9).*

Comment peut-on expliquer d'un point de vue physique que ce modèle de mélange semble représenter de manière plus pertinente le mode de défaillance de nos éprouvettes ? Une défaillance de type risques concurrents semble en effet mieux adaptée à la structure même de notre problème : de multiples micro-fissures étant présentes au sein d'une même éprouvette.

Certaines de ces fissures pourront propager et d'autres non, selon la configuration de leur environnement proche. On pourra même observer des coalescences, c'est-à-dire l'interaction de deux fissures proches (ce qui peut ralentir ou accélérer le phénomène de fissuration). Il n'y a donc pas indépendance entre toutes les fissures d'une éprouvette d'une part. D'autre part, il peut

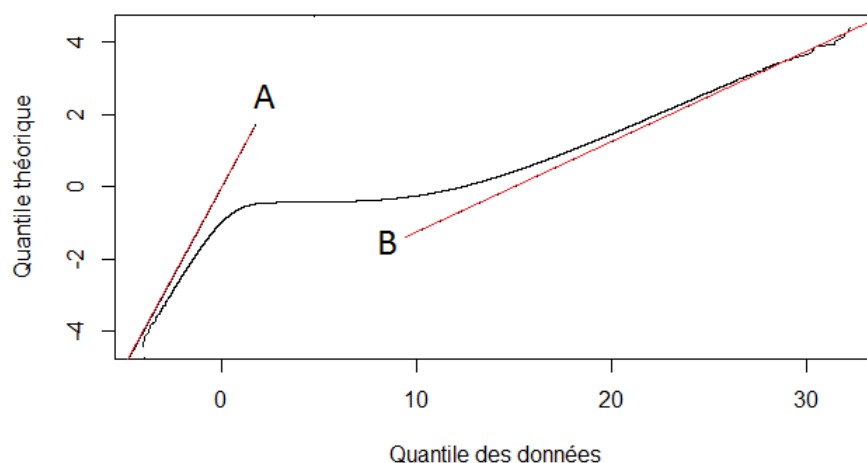


FIGURE 2.9 – quantile plot d’un modèle de mélange de logNormales

arriver, sur certaines éprouvettes, qu’il y ait présence de micro-fissures suffisamment nocives (défaut de la taille du grain par exemple, ou défaut en surface de l’éprouvette) pour que le phénomène menant à la rupture soit beaucoup plus rapide que ce qui est observé dans les cas usuels de fatigue des matériaux (ce phénomène étant aggravé par l’augmentation de température).

La première visualisation macroscopique fait donc apparaître deux modalités de rupture, un mode lent (très fréquent) et un mode rapide (plus rare). Et c’est donc ces deux phénomènes qu’il convient en premier lieu de modéliser.

## 2.5 Conclusion

Les données Snecma que nous étudions sont issues d’essais de fatigue. Nous avons présenté les modélisations classiques basées sur des connaissances mécaniques (défaut, amorçage ou propagation des fissures) ou statistiques (risques concurrents ou modèle de mélanges).

Nous avons constaté que ces modélisations ne permettent pas d’appréhender la variabilité des données Snecma sur l’ensemble des contraintes. Le choix pour notre future modélisation se portera donc sur la construction d’un modèle de mélange au vu des considérations réalisées dans cette section. Toutefois, cette étude ne portait jusque là que sur des considérations

sur les données elles-mêmes. Il nous faut maintenant prendre en compte les phénomènes physiques observés par les chercheurs en matériaux afin d'obtenir un modèle robuste et facilement interprétable.

---



# Chapitre 3

## Endommagement de l'éprouvette

Ce chapitre présente les modes principaux de défaillance dans le cadre de la rupture de fatigue. Il nous permet de comprendre les causes de variabilité des données de durée de vie. La fin du chapitre est consacrée à quelques considérations mathématiques liées à cet endommagement. Il a pour but de présenter succinctement les phénomènes physiques qui conduisent à la rupture de fatigue, notamment ceux qui nous serviront à la formalisation de notre modèle. Pour plus d'information concernant les phénomènes de fatigue et d'amorçages de fissure, le lecteur pourra se référer à Brand et al. [10] ou Bathias et Pineau [6]

### 3.1 Mécanisme d'amorçage

On peut dégager trois phases dans le processus de ruine d'un composant : la phase d'amorçage de la fissure, la phase de propagation de la fissure et la phase de rupture finale. L'amorçage d'une microfissure provient des concentrations de contraintes locales, qu'elles soient dues à la présence d'un défaut (inclusion, accident géométrique de type perçage, ...), d'une interface (joint de grain, surface libre...).

- **Amorçage sur défauts** : amorçage de fissure à partir de défauts extrinsèques au matériau provenant de l'élaboration, des procédés de transformation ou bien de l'usage (inclusions, rayures...).
  - **Amorçage par micro-plasticité cyclique** : amorçage à partir de concentrations de contraintes intrinsèques au matériau (joints de grains, précipités, bandes de glissements persistantes). Les sources de concentrations de contraintes sont très nombreuses et sont générées, ou bien
-

s'accroissent, du fait des microdéformations plastiques cycliques (intrusions-extrusions à l'échelle du grain...).

### 3.1.1 Amorçage sur défauts

Pour les matériaux, les défauts sont des hétérogénéités qui constituent des sites privilégiés pour l'amorçage de fatigue. Leurs caractéristiques tels que la dimension, position, forme et orientation sont conditionnées par le procédé de fabrication. Ces défauts sont généralement dus à des constituants du matériau. Dans le cas de l'Inconel 718, on pourra citer notamment les nitrures et les carbures de niobium visualisables sur la Figure 3.1. Une autre catégorie de défauts est celle constituée par les erreurs humaines lors de la fabrication ou la maintenance de la pièce, notamment les chocs et les rayures. Toutefois cette dernière catégorie est normalement absente des éprouvettes testées.

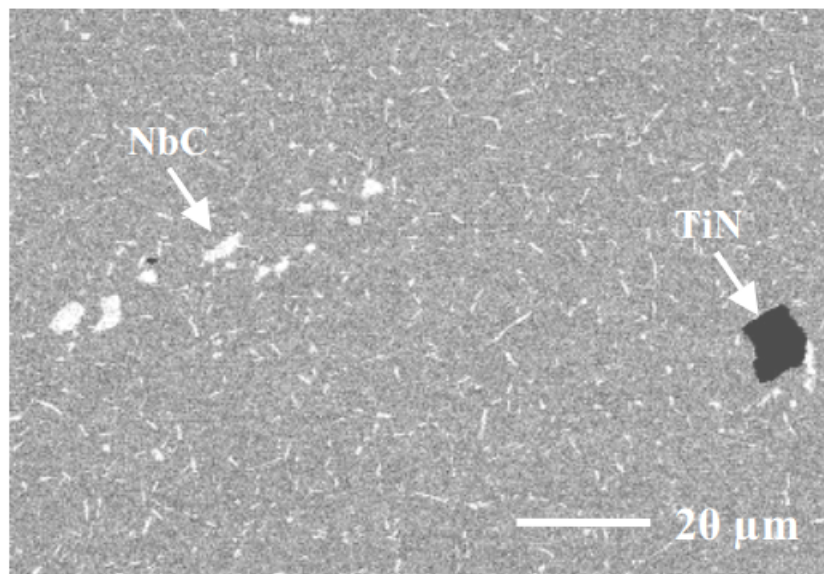


FIGURE 3.1 – Défauts de l'Inconel : nitrures (TiN), carbures de niobium (NbC), Alexandre [4]

Ayant chacun leurs propres caractéristiques chimiques, tous les défauts ne possèdent pas la même nocivité pour la pièce. L'amorçage en fatigue sur un défaut de l'éprouvette dépend donc de l'existence ou non d'un défaut suffisamment nocif dans le volume de cette dernière. D'après Alexandre [4], l'amorçage sur particule est instantané lorsqu'il a lieu. Il se produit au premier quart de cycle en fatigue. La Figure 3.2 représente un amorçage sur

---

défaut subsurfacique avec un cisaillement qui débouche rapidement en surface et une rapide progression de la fissure vers le stade II de la propagation.

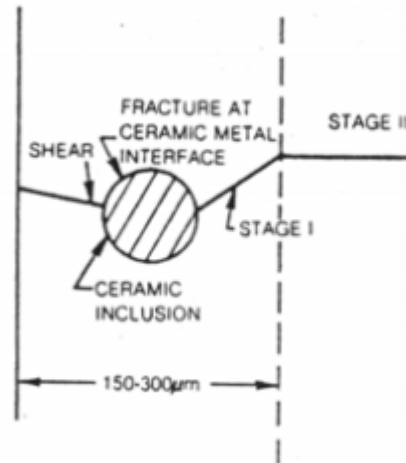


FIGURE 3.2 – Représentation schématique de l'amorçage en coupe sur une inclusion subsurfacique.

Ainsi, bien que les amorçages sur défaut présentent une dispersion de la durée de vie plus faible que les autres amorçages, ils contribuent fortement à la variance totale de la durée de vie en fatigue. En effet, leur durée de vie faible amène de grands écarts avec les autres types d'amorçage et donc une variance inter-groupes (variance issue de l'éloignement entre le groupe des défauts et le groupe de phase I) élevée.

### 3.1.2 Amorçage en phase I (sur micro-plasticité)

Même en absence de défauts dans le volume, des fissures s'initient sous l'effet du chargement cyclique.

Divers mécanismes sont proposés pour expliquer l'amorçage en phase I. Nous présentons ici celui associé aux bandes de glissement. Le lecteur souhaitant un complément d'information pourra s'intéresser au livre de Suresh [44]. A l'échelle du grain, on constate la formation de bandes de glissement dès les premiers cycles de déformation (voir Figure 3.3). Puis, au fur et à mesure, apparaissent des bandes de glissement persistantes traduisant un endommagement irréversible du matériau. À ce stade, la vitesse de propagation de fissure est très lente (de l'ordre de quelques angström :  $1 \text{ \AA} = 10^{-10} \text{ m}$ ). Ainsi, le Stade I, pour les fissures qui ne sont pas liées à des défauts, représente une part importante de la durée de vie totale. De plus, le phénomène est hautement aléatoire, certaines fissures s'arrêteront à la suite de plusieurs barrières

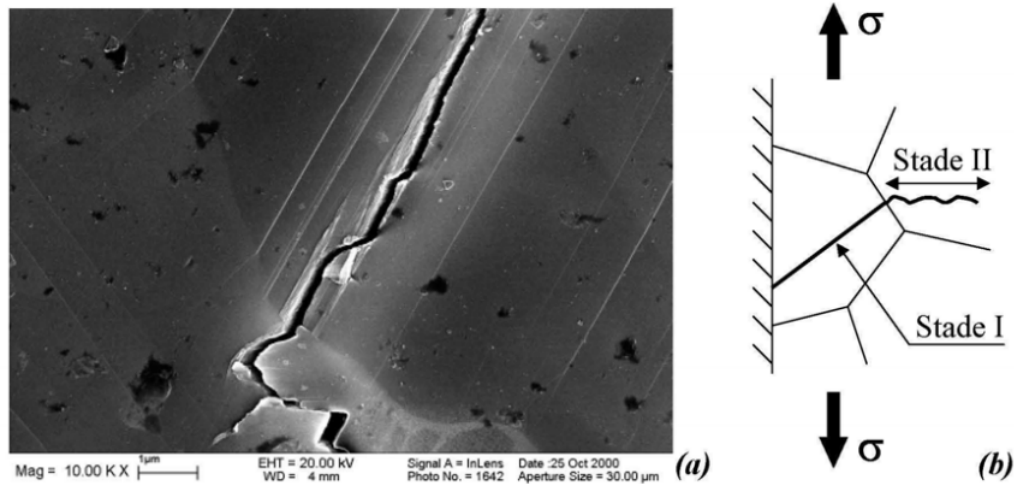


FIGURE 3.3 – a) Amorçage parallèlement aux lignes de glissement du grain. Le Biavant-Guerrier [29] b) Propagation d'une fissure de fatigue

microstructurales tandis que d'autres propageront jusqu'à la rupture. Ce type de rupture présente donc une durée de vie avec grande dispersion.

### 3.2 Propagation de la fissure et rupture

La propagation correspond à la phase II des schémas 3.3 et 3.2. Le passage amorçage/propagation est effectué lorsque la fissure dépasse la taille d'un grain de matériau. On passe alors du mode d'avancement intra-granulaire de la fissure au mode inter-granulaire. Le comportement de la fissure est alors identique quelles que soient les raisons de son amorçage. La modélisation en propagation se fait généralement à l'aide du modèle de Paris (vu au § 2.1). La dispersion de la durée de vie liée à cette propagation est surtout due aux différences entre les tailles de grains de différentes éprouvettes.

La fin de la propagation voit une accélération rapide de la fissure jusqu'à la rupture. Celle-ci ne représente que peu de cycles au vu de la durée de vie d'une éprouvette et n'apporte finalement que peu de dispersion dans les données.

### 3.3 Modèle à deux phases d'amorçage

Quel que soit le mode d'amorçage de l'éprouvette, la durée de vie possède une propagation en phase II qui n'amène que peu de dispersion sur la durée

de vie finale. La dispersion des données est donc particulièrement liée à la phase d'amorçage. Ainsi afin de pouvoir prendre en compte ces deux types d'amorçage, nous pouvons écrire la durée de vie comme la somme d'une phase d'amorçage  $N_i$  et une phase de propagation  $N_p$  :

$$N = N_i + N_p \quad (3.1)$$

Toutefois nous avons vu que les phases d'amorçage sont fondamentalement différentes entre un amorçage sur un défaut ou sur une bande de glissement. Puisque la durée de vie en amorçage sur défaut est négligeable ( $N_i \simeq 0$ ), la durée de vie pourra s'écrire :

$$N = \begin{cases} N_p & \text{si amorçage au premier cycle ;} \\ N_i + N_p & \text{sinon .} \end{cases}$$

Par ailleurs nous pouvons noter que cette expression s'inscrit à la suite du travail d'Alexandre [4]. En effet nous noterons que, bien que nous considérons la propagation comme étant aléatoire, la dispersion des données en propagation est bien plus faible que celle des données présentant un amorçage. Ainsi lorsqu'il y a présence d'un amorçage, la variance est en grande partie expliquée par la durée de cet amorçage. Ce constat est déjà réalisé par Alexandre. La différence d'approche tient principalement aux méthodes de traitement des données. Je ne considère que les données de sortie et d'entrée (la durée de vie et la déformation). Là où Alexandre ajoute une expertise matériau en amenant des paramètres supplémentaires.

Nous avons donc là, la base de la modélisation qui sera mise en place au cours de cette thèse.

## 3.4 Traitement mathématique de la somme

La modélisation retenue contient une somme de deux durées de vie, celle en amorçage et celle en propagation. Nous avons vu au cours de la section 2.4 que des lois lognormales semblent s'adapter à nos données. Or la somme de lognormales n'est pas calculable littéralement et nécessite un traitement particulier que nous présentons dans cette section.

### 3.4.1 Distribution lognormale

Une variable aléatoire  $Z$  suit une loi lognormale de paramètres  $\theta = (\mu, \sigma)$ , lorsque  $Y = \log(Z)$  est une variable aléatoire gaussienne de moyenne  $\mu$  et de variance  $\sigma$ . Sa densité visualisable sur la figure 3.4 est la suivante :

$$\phi(x) = \frac{1}{x\sqrt{2\pi\sigma}} e^{-\frac{(\log(x)-\log(\mu))^2}{2\sigma^2}}, \forall x > 0. \quad (3.2)$$

Ainsi, la loi lognormale est une simple transformation d'une loi normale. Toute une partie des propriétés asymptotiques de la loi normale reste valable dans le cas de la loi lognormale.

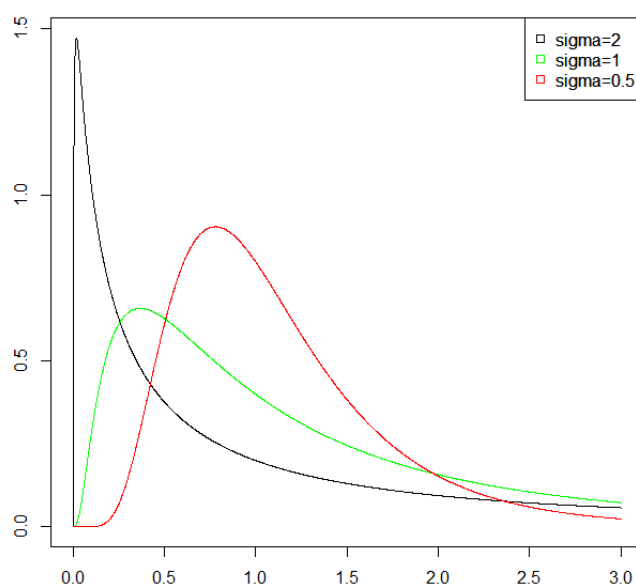


FIGURE 3.4 – Représentation de la loi lognormale pour différents paramètres de forme avec  $\mu = 0$

Cette loi est régulièrement utilisée en fiabilité. En effet, elle possède un sens du point de vue de la dégradation d'un composant mécanique. Considérons une éprouvette sujette à des contraintes et comportant une fissure initiale. Notons  $X_0$  la largeur de celle-ci. Suite à chaque cycle nous mesurons la largeur du défaut, et nous notons donc  $X_1, X_2, \dots, X_n$  les variables aléatoires des  $n$  mesures successives de ces longueurs.

À chaque cycle  $i$ , la largeur du défaut augmente donc de  $X_i - X_{i-1}$ . Supposons de plus qu'une contrainte fait augmenter la fissure de manière proportionnelle à sa taille :

soit  $A_i$  une variable aléatoire, telle que  $(X_i - X_{i-1}) = A_i X_i$ . Au bout de  $n$  cycles on peut donc observer :

$$\sum_{i=1}^n A_i = \sum_{i=1}^n \frac{X_i - X_{i-1}}{X_i}.$$

Maintenant supposons que l'évolution de la fissure est lente, on peut donc utiliser l'approximation suivante :

$$\sum_{i=1}^n A_i = \int_{X_0}^{X_n} \frac{dX}{X},$$

$$\sum_{i=1}^n A_i = \log(X_n) - \log(X_0).$$

En réécrivant cette formule, on peut donc écrire une approximation de  $\log(X_n)$  comme une somme de variables aléatoires  $(A_i)_{i=1\dots n}$  :

$$\log(X_n) = \sum_{i=1}^n A_i + \log(X_0).$$

On peut enfin utiliser le théorème central-limite :

$$\frac{\sum_{i=1}^n A_i - \mu}{\sqrt{n\sigma}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1).$$

Ainsi le logarithme de la largeur de  $X_n$  est asymptotiquement une loi normale. Cette loi est donc naturellement la loi de distribution des défaillances d'usure mécanique simple.

### 3.4.2 Distribution de somme de lognormales

On a vu précédemment qu'une des possibilités de rupture est la somme d'un amorçage long et de propagation. Or nous avons vu dans la section 2.4 que ces deux durées de vies semblent suivre des lois log-normales. Toutefois le traitement d'une somme de lois lognormales est complexe.

La complexité de la somme d'une loi de log-normale vient du fait que son expression sous forme de produit de convolution possède une expression ne conduisant pas à une formule explicite :

$$f_{N_i+N_p}(n) = [f_{N_i} * f_{N_p}](n) = \int_0^n f_{N_i}(n-x) f_{N_p}(x) dx. \quad (3.3)$$

L'utilisation de cette expression passe donc nécessairement par une approximation. On retrouve cette problématique liée à la somme de lognormales dans les télécommunications, la finance et l'ingénierie du trafic. Tandis que l'expression de la somme de lognormales est inconnue, plusieurs méthodes analytiques d'approximation ont été proposées. (cf. Fenton [20])

### 3.4.3 Approximation par une autre distribution

La première approche de ce type d'approximation fut réalisée en 1960 par Fenton [20]. Nous allons la décrire ici.

On se place dans le cas de deux variables aléatoires lognormales, la loi de la somme  $S = \sum_{i=1}^2 N_i$  peut être approchée par une loi lognormale unique :  $Z \sim \mathcal{LN}(\mu_Z, \sigma_Z)$ .

L'espérance et la variance de  $Z$  sont obtenues par égalité des premiers moments  $m_1$  et  $m_2$  de la somme  $S$ . Cette méthode (dite des moments) permet d'obtenir facilement et rapidement des valeurs pour les paramètres de la lognormale qui approche la loi de la somme :

$$m_1 = \mathbb{E}(S) = e^{\mu_Z + \sigma_Z^2/2} = \sum_{i=1}^2 e^{\mu_{N_i} + \sigma_{N_i}^2/2},$$

$$m_2 = \mathbb{E}(S^2) = e^{2\mu_Z + 2\sigma_Z^2} = \sum_{i=1}^2 e^{2\mu_{N_i} + 2\sigma_{N_i}^2} + 2e^{\mu_{N_1} + \mu_{N_2}} e^{\frac{\sigma_{N_1}^2 + \sigma_{N_2}^2}{2}}.$$

En résolvant ce système on peut obtenir  $\mu_Z$  et  $\sigma_Z$  :

$$\mu_Z = 2 \log(m_1) - \frac{1}{2} \log(m_2),$$

$$\sigma_Z^2 = \log(m_2) - 2 \log(m_1).$$

L'inconvénient de cette méthode est que, plus l'écart entre les variances des lognormales composant la somme est importante, moins l'approximation est bonne (voir Schwartz et Yeh [41]).

Depuis, d'autres méthodes ont été proposées afin d'en atténuer les faiblesses. Certains (Wu et al. [49]) essaient d'améliorer l'approximation en proposant d'utiliser un plus grand nombre de moments afin de mieux approcher la distribution de la somme, notamment dans les queues. D'autres préfèrent approcher la loi de la somme par des lois de type log-skew-normal (Li 2008 [30]), modified-power-lognormal (Szyzkowicz et Yanikomeroglu [45]). Tous se basent sur une égalité des moments entre la somme et la distribution voulue.

Le premier problème de ces modélisations est qu'elles sont validées sur un grand nombre de loi lognormales, ainsi leur comportement est moins adéquat lorsque l'on ne travaille que sur deux composantes comme nous le faisons. De plus, toutes ces approximations n'approchent bien la loi que sur un domaine



restreint. Ainsi, la loi de la somme ne peut être intégralement représentée par une de ces approximations. La Figure 3.5 montre la loi de la somme et ses approximations. On constate que si la log-skew-normal semble être la distribution la plus proche de la loi réelle, elle fait toutefois défaut pour la queue basse de la distribution. Or c'est dans cette partie que nous avons besoin d'être le plus précis puisque ce sont ces queues basses qui fournissent le quantile à 0,1% du modèle.

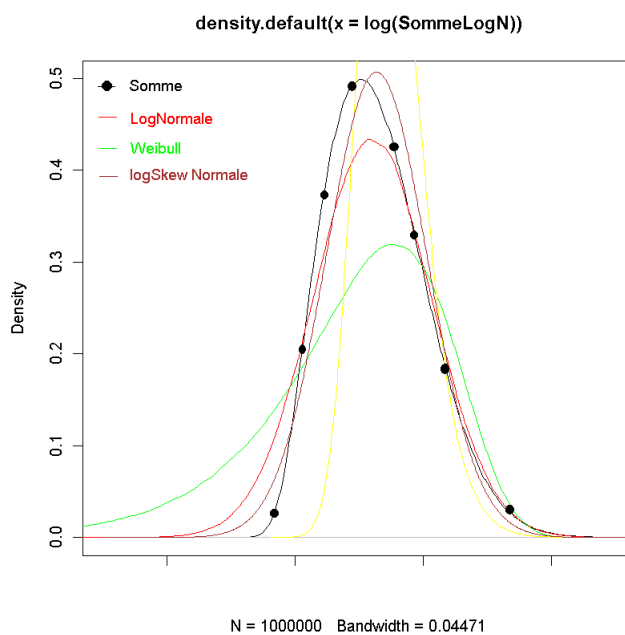


FIGURE 3.5 – Représentation de la loi de la somme et différentes approximations par une autre loi

Ainsi ces méthodes ne seront pas retenues.

### 3.4.4 Approximations numériques

Cette section présente les techniques usuelles d'approche numérique de calcul d'intégrales.

#### Méthode de quadrature de Gauss

La méthode de quadrature de Gauss (Abramowitz et Stegun [1]) consiste à calculer de façon approchée une intégrale de la forme :

$$I = \int_{\mathcal{D}} f(x)W(x)dx,$$

où  $W : \mathcal{D} \rightarrow \mathbb{R}_+$  est une fonction de pondération, et  $f : \mathbb{R} \rightarrow \mathbb{R}$  est une fonction régulière sur  $\mathcal{D}$ .

Le principe de la méthode de quadrature de Gauss est le suivant : soit  $N$  un entier, on peut trouver une famille de  $N$  réels positifs  $\omega_i$  appelés poids et  $N$  réels  $x_i$  appelés points d'intégration tels que l'intégrale  $I$  soit approchée par l'expression suivante :

$$I \simeq \sum_{i=1}^N \omega_i f(x_i).$$

Pour déterminer les points  $x_i$  et les poids  $\omega_i$  qui leur sont associés, on prend ces points comme les racines de polynômes orthogonaux.

*Remarque : Les polynômes orthogonaux sont choisis en fonction du domaine d'intégration.*

Notre intégrale étant définie sur l'intervalle  $[0, n]$  borné, les polynômes choisis seront ceux de Legendre définis comme suit :

$$P_N(x) = \frac{1}{2^N} \sum_{k=1}^N \binom{N}{k}^2 (x-1)^{N-k} (x+1)^k.$$

Les poids sont quant à eux donnés par la relation suivante :

$$\omega_i = \frac{-2}{(N+1)P'_N(x_i)P_{N+1}(x)}.$$

La fonction de poids étant :  $W(x) = 1$ .

L'utilisation d'une méthode de quadrature de Gauss permet, la plupart du temps, de réduire le nombre d'appels au modèle par rapport à une méthode de Monte-Carlo. Toutefois dans notre cas, un phénomène rend l'estimation très mauvaise lorsque nous ne sommes pas en présence d'un très grand nombre de points  $x_i$  :

La Figure 3.6 représente une partie de la fonction à intégrer. La fonction est quasiment plate sur une grande partie de la courbe sauf sur un petit domaine où elle prend de grandes valeurs. Ainsi, une bonne approximation de l'intégrale de cette courbe nécessite de nombreux points, puisqu'il faut au moins assurer la présence d'un ou deux points dans cette partie piquée de la courbe. Afin de diminuer le nombre de points utilisés pour réaliser le calcul, nous proposons de réaliser une intégration Monte-Carlo. Celle-ci permettra d'aller chercher plus de points au sein de ce pic et donc permettra d'améliorer la qualité de l'approximation.

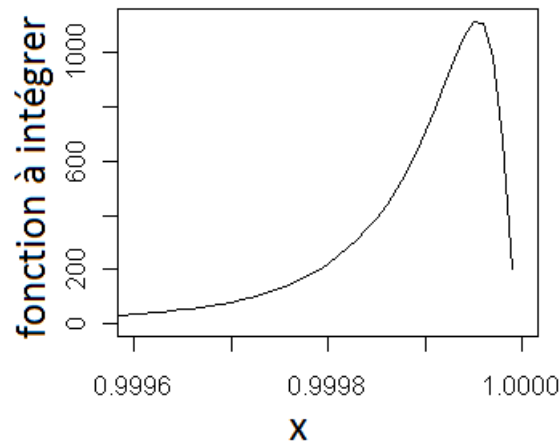


FIGURE 3.6 – Zoom sur la fonction à intégrer

### Méthode de Monte-Carlo

Les méthodes de Monte-Carlo regroupent un grand nombre de méthodes utilisant des processus aléatoires (et des simulations) afin de réaliser des approximations numériques. Dans notre cas, nous les utiliserons pour le calcul de l'intégrale (3.3).

Pour estimer une intégrale, il s'agit de l'exprimer dans un premier temps en tant qu'espérance mathématique. Soit  $f$  la densité d'une variable aléatoire  $X$ . On peut écrire l'espérance suivante :

$$I = \int g(x)f_X(x)dx = E[g(X)].$$

L'idée est donc de simuler  $(x_1, \dots, x_N)$   $N$  points suivant la loi de  $X$ , et de calculer un estimateur de  $E(g(X))$  à partir de cet échantillon. Soit,

$$\hat{g}_N = \frac{1}{N} \sum_{i=1}^N g(x_i),$$

un estimateur sans biais de la quantité  $E(g(X))$ . Cet estimateur est appelé estimateur de Monte-Carlo de I.

Dans notre cas, comme l'intégrale est le produit de convolution de deux densités on peut écrire l'intégrale comme l'espérance de l'une ou l'autre des densités :

$$f_{N_i+N_p}(n) = \int_0^n f_{N_i}(n-x)f_{N_p}(x)dx \quad (3.4)$$

$$= \mathbf{E}_{N_p}[f_{N_i}(n-N_p)^+] \quad (3.5)$$

$$= \mathbf{E}_{N_i}[f_{N_p}(n-N_i)^+]. \quad (3.6)$$

Elle est donc estimée par échantillonnage préférentiel (afin d'obtenir la plus petite variance possible) en utilisant soit (4.9), soit (4.10), suivant que  $\hat{\sigma}(N_p) < \hat{\sigma}(N_i)$  ou non. Sans perte de généralité, nous considérerons que  $\hat{\sigma}(N_p) < \hat{\sigma}(N_i)$ . La procédure est alors la suivante :

1.  $b$  simulations indépendantes de  $f_{N_p}$  sont réalisées :  $n_{p_j}$ ,  $j = 1, \dots, b$ .  
En pratique,  $b = 1000$  semble offrir des résultats satisfaisants.
2. D'après (4.9),  $f_{N_i+N_p}(n)$  est estimée par

$$\hat{f}_{N_i+N_p}(n) = \frac{1}{b} \sum_{j=1}^b f_{n_i}(n-n_{p_j})^+.$$

### 3.4.5 Approximation des queues de distribution

Un des autres points importants à étudier est le comportement de cette somme dans les queues de la distribution.

La loi lognormale est une loi de distribution à queue lourde, plus exactement elle est sous-exponentielle. Ce nom provient de la propriété de décroissance des queues plus lentes que celle d'une exponentielle. Cela implique que de grandes valeurs peuvent apparaître dans un échantillon avec une probabilité non négligeable.

**Définition : Distribution sous-exponentielle**

Soient  $(X_i)_{i \in \mathbb{N}}$  des variables i.i.d. positives avec une fonction de répartition  $F$ , telle que  $F(x) < 1, \forall x > 0$ .

Notons,

$$\bar{F}(x) = 1 - F(x), x > 0,$$

la queue de  $F$  et

$$\bar{F}^{n*} = 1 - F^{n*} = P(X_1 + \dots + X_n > x),$$

la queue du  $n$ -produit de convolution de  $F$ .

$F$  est une fonction de répartition sous exponentielle si une des deux conditions équivalentes suivante est vérifiée :

1.  $\lim_{x \rightarrow \infty} \frac{\bar{F}^{n*}(x)}{\bar{F}(x)} = n, \forall n > 1$
2.  $\lim_{x \rightarrow \infty} \frac{P(X_1 + \dots + X_n > x)}{P(\max(X_1, \dots, X_n) > x)} = 1, \forall n > 1.$

*Remarque : La définition (1.) vient de Chistyakov [15]. Tandis que l'équivalence entre les deux conditions a été démontrée par Embrechts et Goldie [19].*

Cette définition nous permet d'avoir une interprétation simple du comportement des queues d'une loi sous-exponentielle. Si nous prenons la somme de  $n$  distributions sous-exponentielle i.i.d., cette somme dépasse une valeur extrême si et seulement si un des termes de la somme le dépasse.

### Equivalence des queues d'une somme de deux lognormales

Pour la suite, pour simplifier, nous considérerons que  $\sigma_2 > \sigma_1$ . Ainsi nous pouvons écrire une approximation des queues de distribution pour une somme de deux lognormales :

Soit

$$N_1 \sim \mathcal{LN}(\mu_1, \sigma_1), N_2 \sim \mathcal{LN}(\mu_2, \sigma_2),$$

alors

$$\begin{aligned} \mathbb{P}(N > c) &= \mathbb{P}(N_1 + N_2 > c) \underset{c \rightarrow \infty}{\simeq} P(N_2 > c) \\ &\underset{c \rightarrow \infty}{\simeq} \frac{\sigma_2 e^{-\frac{1}{2} \frac{(\log(c) - \mu_2)^2}{\sigma_2^2}}}{\sqrt{2\pi}(\log(c) - \mu_2)}. \end{aligned}$$

## 3.5 Conclusion

Ce chapitre, afin d'améliorer une modélisation purement statistique par modèle de mélange, tente d'expliquer la survenue de la bi-modalité dans nos données. Ainsi, elle peut s'expliquer par un phénomène d'amorçage rapide ou non en fonction du type de défauts présents dans l'éprouvette. Ceci donne lieu à deux comportements : les éprouvettes qui rompent en « propagation pure », et celles à durée de vie plus élevée qui voient d'abord une phase d'amorçage longue, suivie d'une propagation similaire aux autres éprouvettes. Nous voyons ici apparaître l'importance de séparer la durée de vie totale de nos éprouvettes en deux, la partie propagatoire et la partie amorçage. La deuxième partie de ce chapitre est donc consacrée aux différentes méthodes statistiques qui permettent de traiter ce type de somme de distribution, afin de pouvoir les intégrer dans le modèle complet qui sera proposé au chapitre suivant. Notre modèle utilise différents paramètres dont l'information amorçage lent ou rapide. Or cette dernière n'est pas renseignée dans la base Snecma.

Nous allons donc, au cours du chapitre suivant, mettre en place l'estimation des différents paramètres.

---

## Chapitre 4

# Un modèle de mélange Amorçage-Propagation

Ce chapitre présente un nouveau modèle de courbe  $S - N$  qui a pour principal objectif de prendre en compte la bi-modalité des données observées et montrées au chapitre précédent. Il a aussi l'avantage d'être utilisable sans données fractographiques, ce qui le rend particulièrement intéressant pour des bases de données industrielles telles que celles utilisées à la Snecma. L'analyse fractographique n'étant pas systématiquement effectuée compte tenu de son coût. Il a fait l'objet d'un article soumis à *l'International Journal of Fatigue*.

Nous proposons un modèle basé sur la mécanique de la rupture. Il exploite le fait que la fatigue peut être vue comme la somme d'une période d'amorçage et d'une période de propagation. La phase d'amorçage,  $N_i$ , est définie comme le nombre nécessaire de cycles afin de former une petite fissure visible (de la taille du grain du matériau). La durée de vie en propagation,  $N_p$ , est le nombre de cycles requis pour étendre cette fissure jusqu'à la taille critique de rupture de l'éprouvette. Un test de fatigue peut donc s'écrire  $N = N_p$  lorsque la fissure apparaît dès le premier chargement en contrainte de l'éprouvette, ou sinon  $N = N_p + N_i$ . Ce comportement nous amène au modèle de mélange suivant :

$$f_N = \pi f_{N_p} + (1 - \pi) f_{N_i + N_p}, \quad (4.1)$$

$\pi$  étant la probabilité que la fissure se forme au premier chargement de l'éprouvette. Toutefois dans notre situation, l'information n'est pas présente, nous sommes donc en présence d'un problème de données manquantes.

---

## 4.1 Proportion constante

Nous supposons ici que  $\pi(S)$  est une constante (et ne dépend donc pas de  $S$ ). Comme vu précédemment la fatigue peut s'écrire :

$$N = \begin{cases} N_p & \text{si } Z = 1; \\ N_i + N_p & \text{si } Z = 0, \end{cases}$$

où  $Z$  est le label pour un amorçage au premier chargement ( $Z = 1$  si la fissure démarre au premier cycle, 0 sinon). Nous supposons des distributions lognormales pour  $N_i$  et  $N_p$ , ce qui semble raisonnable au vu de nos données (cf. §2.6).

$$f_{N_i}(n, s) = \frac{1}{n \sigma_i \sqrt{2\pi}} \exp\left(-\frac{[\ln(n) - (a_i s + b_i)]^2}{2\sigma_i^2}\right); \quad (4.2)$$

$$f_{N_p}(n, s) = \frac{1}{n \sigma_p \sqrt{2\pi}} \exp\left(-\frac{[\ln(n) - (a_p s + b_p)]^2}{2\sigma_p^2}\right), \quad (4.3)$$

$s$  étant le niveau de déformation du test. La durée de vie  $N$ , peut donc s'écrire comme un mélange entre un mode de propagation pure  $N_p$ , et un mode composé d'amorçage et propagation  $N_i + N_p$  :

$$f_N = \pi f_{N_p} + (1 - \pi) f_{N_i + N_p}. \quad (4.4)$$

Le terme  $N_i + N_p$  représente le « comportement normal » en fatigue, tandis que le terme  $N_p$  correspond à un comportement moins fréquent de « propagation pure ». La probabilité  $\pi$  représente la probabilité d'être dans le mode de propagation pure et sera considérée comme constante dans cette première partie.

*Remarque : L'identifiabilité du modèle est démontrée dans l'annexe A. Cette propriété est importante puisqu'elle assure qu'à une distribution de notre modèle, correspond un seul jeu de paramètres.*

### 4.1.1 Estimation des paramètres du modèle

Comme nous n'avons pas d'information sur l'amorçage des éprouvettes, notre modèle (4.4) possède une structure de données manquantes. Afin d'en estimer les paramètres, nous utiliserons l'algorithme d'Espérance-Maximisation (EM, cf. Dempster et al. [17]). L'étape de maximisation de la log-Vraisemblance des données est ici complexe puisqu'elle implique une fonction non-linéaire des paramètres. La résolution nécessite alors un calcul d'intégrale à l'aide



de simulations de Monte-Carlo, en plus de l'algorithme de Newton-Raphson utilisé pour la maximisation.

Soit  $\theta = (\theta_p, \theta_i, \pi)$  le vecteur des paramètres du modèle (4.4), avec  $\theta_p = (a_p, b_p, \sigma_p)$  et  $\theta_i = (a_i, b_i, \sigma_i)$ .

L'algorithme EM est un algorithme en deux étapes ayant pour but de maximiser la vraisemblance des données observées :

$$L(N, S; \theta) = \prod_k^m f_{(N,S)}(n_k, s_k; \theta). \quad (4.5)$$

On peut écrire :

$$L(N, S; \theta) = L(N, S, Z; \theta) - \sum_i g(z_i | n_i, s_i; \theta),$$

où  $g(z_i | n_i, s_i; \theta)$  est la probabilité d'avoir  $z_i = 1$ , sachant les données  $(n_i, s_i)$  et le paramètre  $\theta$ . Où

$$L(N, S, Z; \theta) = \prod_k^m f_{(N,S,Z)}(n_k, s_k, z_k; \theta), \quad (4.6)$$

est la vraisemblance des données complétées. Avec

$$f_{(N,S,Z)}(n, s, z; \theta) = (\pi(s)[f_{N_i}(n, s)])^z \times ((1 - \pi(s))[f_{N_i+N_p}(n, s)])^{(1-z)}.$$

*Remarque : Z est le label inconnu lié à l'origine de la fissure : Z = 1 si la fissure a lieu au premier chargement, 0 sinon.*

Comme la variable  $Z$  n'est pas observée, la vraisemblance des données complétées ne peut être calculée directement. L'étape « Expectation » consiste à calculer l'espérance (sur  $Z$ ) de cette vraisemblance à l'aide de valeurs courantes pour les paramètres  $\theta$ . L'étape « Maximisation » revient à maximiser (en  $\theta$ ) cette espérance. Ces deux étapes sont ensuite réitérées jusqu'à convergence de l'algorithme. On démontre ensuite que le maximum atteint est bien celui du maximum de vraisemblance des données observées (4.5).

### Algorithme EM 1

En partant d'un vecteur de données initiales  $\theta^{(0)} = (\theta_p^{(0)}, \theta_i^{(0)}; \pi^{(0)})$ , l'algorithme itère les étapes E et M suivantes :

1. Étape E : Calcul de  $Q(\theta|\theta^{(j)})$  :

$$\begin{aligned} Q(\theta|\theta^{(j)}) &= \mathbf{E} [\ln L(N, S, Z; \theta) | (Z|N; \theta^{(j)})] \\ &= \sum_k^m [\hat{t}_k \ln(f_{N_p}(n_k, s_k; \theta_p)) \\ &\quad + (1 - \hat{t}_k) \ln(f_{N_i+N_p}(n_k, s_k; \theta))]; \end{aligned} \quad (4.7)$$

$\hat{t}_k$  étant la distribution conditionnelle de  $Z$  :

$$\begin{aligned} \hat{t}_k &= \mathbf{E}[Z|N, S; \theta^{(j)}] \\ &= \mathbf{P}(Z = 1 | N = n_k, S = s_k; \theta = \theta^{(j)}) \\ &= \frac{\pi^{(r)}(s_k) f_{N_p}(n_k; s_k, \theta_p^{(r)})}{\sum_{l=(1,2)} \pi_l(s_k) \phi_l(n_k; s_k, \theta^{(r)})}; \end{aligned}$$

où  $\phi_1$  remplace  $f_{N_p}$  et  $\phi_2$  remplace  $f_{N_i} * f_{N_p} = f_{N_i+N_p}$ .

2. Étape M : maximisation de  $Q(\theta|\theta^{(j)})$

$$\hat{\theta}^{(j+1)} = \arg \max_{\theta} Q(\theta|\theta^{(j)}).$$

Ce qui peut être décomposé en deux maximisations :

$$\begin{aligned} & - \max_{\pi} \sum_{k=1}^N \hat{t}_k \ln(\pi) + (1 - \hat{t}_k) \ln(1 - \pi); \\ & - \max_{\theta} \sum_{k=1}^N \hat{t}_k \ln(f_{N_p}(n_k, s_k, \theta_p)) + (1 - \hat{t}_k) \ln(f_{N_i} * f_{N_p}(n_k, s_k, \theta)). \end{aligned}$$

Ce qui nous conduit à trouver les zéros des dérivées suivantes :

$$\left\{ \begin{array}{l} \sum_{i=1}^N t_{i2} \left( \frac{\log(n_i) - \mu_p}{\sigma_p^2} + \frac{\int_0^{n_i} \frac{\log(l) - \mu_p}{l(n_i-l)\sigma_p^3\sigma_a} A(n_i, l, \mu_a, \mu_p, \sigma_a, \sigma_p) dl}{\int_0^{n_i} \frac{A(n_i, l, \mu_a, \mu_p, \sigma_a, \sigma_p)}{l(n_i-l)\sigma_p\sigma_a} dl} \right) \\ \sum_{i=1}^N S_i t_{i2} \left( \frac{\log(n_i) - \mu_p}{\sigma_p^2} + \frac{\int_0^{n_i} \frac{\log(l) - \mu_p}{l(n_i-l)\sigma_p^3\sigma_a} A(n_i, l, \mu_a, \mu_p, \sigma_a, \sigma_p) dl}{\int_0^{n_i} \frac{A(n_i, l, \mu_a, \mu_p, \sigma_a, \sigma_p)}{l(n_i-l)\sigma_p\sigma_a} dl} \right) \\ \sum_{i=1}^N t_{i2} \left( \frac{-1}{\sigma_p} + \frac{(\log(n_i) - \mu)^2}{\sigma_p^3} + \frac{\int_0^{n_i} \frac{A(n_i, l, \mu_a, \mu_p, \sigma_a, \sigma_p)}{l(n_i-l)\sigma_a\sigma_p} \left( -\frac{1}{\sigma_p} + \frac{(\log(l) - \mu_p)^2}{\sigma_p^3} \right) dl}{\int_0^{n_i} \frac{A(n_i, l, \mu_a, \mu_p, \sigma_a, \sigma_p)}{l(n_i-l)\sigma_p\sigma_a} dl} \right) \end{array} \right.$$

$$\text{avec } A(n_i, l, \mu_a, \mu_p, \sigma_a, \sigma_p) = e^{-\frac{(\log(n_i-l) - \mu_a)^2}{2\sigma_a^2}} e^{-\frac{(\log(l) - \mu_p)^2}{2\sigma_p^2}}$$

*Remarque, nous donnons ici uniquement la dérivée par rapport à  $\theta_p$ , Les équations pour  $\theta_a$  s'obtiennent de manière similaire.*

Les zéros sont calculés à l'aide d'un algorithme 'BFGS' dérivé de celui de Newton-Raphson voir Annexe B. Cet algorithme a l'avantage de ne

pas nécessiter de calcul pour la dérivée seconde nécessaire dans l'algorithme usuel de Newton-Raphson. Ici, cette dérivée est très lourde à calculer de manière numérique.

La sequence  $\theta^{(1)}, \theta^{(2)}, \dots$  générée par EM converge vers un maximum local de la **log-vraisemblance des données observées** sous des conditions assez générales (cf. Dempster et al. [17])

La maximisation de  $Q(\theta|\theta^{(j)})$  nécessite, en plus de l'utilisation d'un algorithme de Newton-Raphson, l'utilisation d'un algorithme de Monte-Carlo afin d'évaluer la densité de  $f_{N_i+N_p}$ . Cet algorithme est décrit ci-dessous.

### Produit de convolution

Le produit de convolution, utilisé en (4.7) ne mène pas à une expression explicite :

$$f_{N_i+N_p}(n) = f_{N_i * f_{N_p}}(n) = \int_0^n f_{N_i}(n-x)f_{N_p}(x)dx \quad (4.8)$$

$$= \mathbb{E}_{N_p}[f_{N_i}(n - N_p)^+] \quad (4.9)$$

$$= \mathbb{E}_{N_i}[f_{N_p}(n - N_i)^+]. \quad (4.10)$$

La densité est donc approchée par échantillonnage préférentiel en utilisant soit l'espérance (4.9), soit (4.10), suivant que  $\hat{\sigma}(N_p) < \hat{\sigma}(N_i)$  ou non. Sans perte de généralité, nous considérerons que  $\hat{\sigma}(N_p) < \hat{\sigma}(N_i)$ . La procédure est alors la suivante :

1.  $b$  simulations indépendantes de  $f_{N_p}$  sont réalisées :  $n_{p_j}$ ,  $j = 1, \dots, b$ .  
En pratique,  $b = 1000$  semble offrir des résultats satisfaisants.
2. D'après (4.9),  $f_{N_i+N_p}(n)$  est estimée par :

$$\hat{f}_{N_i+N_p}(n) = \frac{1}{b} \sum_{j=1}^b f_{n_i}(n - n_{p_j})^+.$$

### Initialisation de l'algorithme

Les résultats donnés par EM peuvent fortement dépendre des conditions d'initialisation de l'algorithme. Toutefois pour notre modèle, l'algorithme semble assez stable. Nous recommandons cependant l'utilisation des valeurs suivantes comme conditions initiales :

- $\pi^{(0)} = 0, 5$  ;

- $\theta^{(0)} = (\theta_p^{(0)}, \theta_i^{(0)})$  sont issues d'une première régression par morceaux des log-durées de vie (cf DeSarbo[18]).

Pour les régions à faible durée de vie  $N = N_p$ , tandis que pour les fortes durées de vie  $N \simeq N_i$ . La régression par classe apporte donc une première estimation des paramètres du modèle assez proche des paramètres finaux. Ce qui permet à notre algorithme de converger plus rapidement (voir §4.3.1 pour plus de détails).

### Cas des données censurées

Les données de fatigue présentent souvent des censures à droite : l'essai est stoppé après un grand nombre de cycles (de l'ordre de  $10^7$ ). La vraisemblance inclut, dès lors, la probabilité suivante :  $\mathbb{P}(N > c)$ , qui est la probabilité que la censure soit dépassée. Toutefois comme nous n'avons pas accès à la densité de la somme de lognormales, cette quantité est difficile à calculer. En effet cela requiert l'approximation numérique de deux intégrales. La densité du mélange en un point censuré est exprimée ci-dessous. Nous notons  $c$  la valeur de la censure :

$$\mathbb{P}(N > c) = \pi \int_c^\infty f_{N_i}(l; \theta, s) dl + (1 - \pi) \int_c^\infty f_{N_i+N_p}(l; \theta, s) dl,$$

où  $f_{N_i+N_p}$  ne possède pas de formule explicite.

Deux solutions s'offrent à nous :

- a. utiliser une approximation asymptotique (cf. Asmussen [5]), nous avons en effet :

$$\mathbb{P}(N > c) \underset{c \rightarrow \infty}{\sim} \frac{\pi_2 \sigma_i e^{-\frac{1}{2} \frac{(\log(c) - \mu_i)^2}{\sigma_i^2}}}{\sqrt{2\pi}(\log(c) - \mu_i)}.$$

Mais pour les faibles déformations, cette asymptote n'est pas réaliste puisque la censure peut survenir bien avant la queue de distribution.

- b. Simuler les données au-delà de la censure  $c$ . L'algorithme EM est alors remplacé par un algorithme SEM qui ajoute une étape stochastique (cf. Celeux et Dieboltz [13]). Toutefois ceci peut prendre un temps assez long lorsque la censure se situe dans la queue de distribution.

Afin d'outrepasser ces problèmes, nous proposons l'heuristique suivante qui donne de bons résultats pour la mise en application.

- Les censures seront considérées comme appartenant à la modalité  $f_{N_i+N_p}(n)$ , puisque c'est l'amorçage qui amène de longues durées de vie. On prendra donc  $t_{k1} = 0$  pour les données censurées.
- Avant chaque étape de maximisation, simuler jusqu'à 1000 fois les deux log-normales et les sommer :

- la première valeur des 1000 qui dépassera la censure sera reprise dans l'étape de maximisation et considérée comme n'importe quel autre point ;
- si cette censure n'est pas dépassée, nous considèrerons que nous sommes dans la queue de la distribution et que l'approximation asymptotique peut donc être utilisée.

Dans la pratique, la phase d'utilisation de l'approximation n'est réalisée que dans les premiers pas de l'algorithme, si l'initialisation des paramètres est trop éloignée de l'optimalité. Sinon, il peut être rare de dépasser la censure, celle-ci se trouvant très loin dans les queues de la distribution. Au fur et à mesure de la convergence des estimateurs, les censures ne dépassent plus la moyenne des données que par deux fois leur écart-type. Ainsi, avec les paramètres finaux sur données réelles, la simulation d'un point est retenue avec une probabilité de  $10^{-5}$ .

#### 4.1.2 Résultats sur données réelles

paramètre	$\pi$	$a_i$	$b_i$	$\sigma_i$	$a_p$	$b_p$	$\sigma_p$
valeur obtenue	0.1	20	-25	1	10	-5	0.2

TABLE 4.1 – Paramètres estimés à 450°C, ordre de grandeurs

Nous nous concentrons ici sur les données réelles. La Table 4.1 présente les estimateurs du modèle (4.4). Les comparatifs sur les données simulées ainsi que les différentes expérimentations afin de calibrer l'algorithme sont donnés en fin de chapitre.

La Figure 4.1 représente le « quantile plot » de nos données avec en ligne continue la projection de ce que donne notre modèle à chaque déformation. Si la partie amorçage semble correctement retrouvée sur ce graphique, on constate l'existence d'un saut aux basses déformations là où il n'existe plus sur les données réelles. Par ailleurs, aux fortes déformation le modèles semble loin des données réelles. Ainsi l'hypothèse d'un  $\pi$  constant semble être une hypothèse forte qui n'est pas vérifiée par les données. Par la suite, il nous faudra donc l'adapter pour pouvoir prendre en compte les variations de proportion de chaque mode en fonction de la déformation. Autrement dit nous devons changer  $\pi$  en  $\pi(S)$ .

Cela implique que l'hypothèse que nous réalisons maintenant, est que la rupture au premier cycle ne dépend pas uniquement de la présence d'un défaut suffisamment gros dans le volume de l'éprouvette. Elle dépend aussi

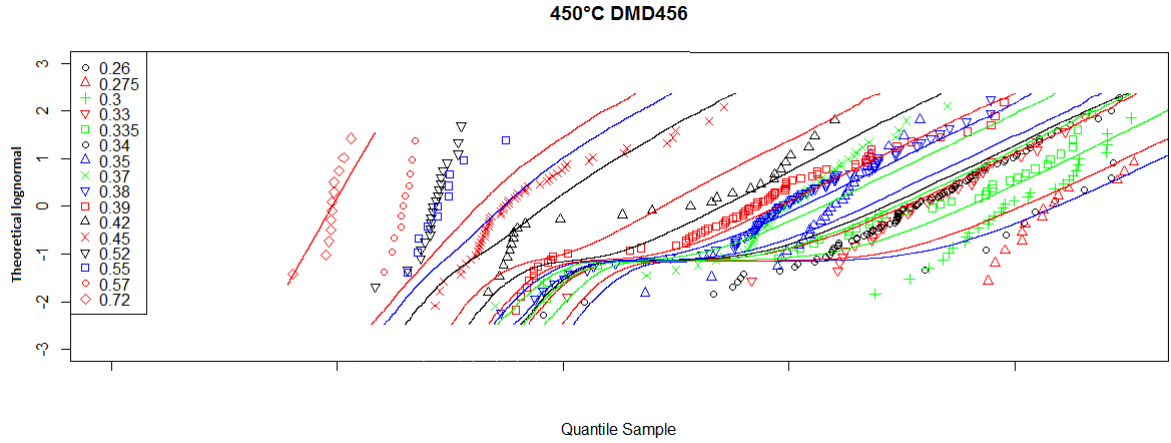


FIGURE 4.1 – quantile plot des données et du modèle avec  $\pi$  constant

de l'énergie d'activation de ce défaut. Un défaut nocif amorcera rapidement aux faibles déformations, là où un défaut plus « doux » nécessitera une déformation plus importante pour amorcer.

## 4.2 Proportion variable

Nous venons de voir qu'afin d'affiner notre modélisation, il est indispensable de faire dépendre  $\pi$  de la contrainte de l'essai. Le modèle s'écrit alors :

$$f_N = \pi(S)f_{N_p} + (1 - \pi(S))f_{N_i+N_p}. \quad (4.11)$$

Toutefois, il reste de multiples possibilités pour la construction d'un  $\pi$  variable. Nous nous concentrerons ici sur celles qui semblent les plus intéressantes dans notre cas, à savoir un  $\pi(S)$  affine par morceau et un  $\pi(S)$  logistique.

La définition de  $\pi(S)$  sera donnée dans chaque sous-section.

### 4.2.1 Proportion affine par morceaux

L'idée est ici qu'il existe une contrainte en deçà de laquelle le mode purement propagatoire n'existe plus, puisque l'énergie fournie n'est plus suffisante pour amorcer la fissure dès le premier chargement de l'éprouvette. Nous reprenons donc le modèle (4.11) en introduisant un  $\pi(S)$  constant par morceau :

$$\pi(S) = \begin{cases} 0 & \text{si } S < s_0; \\ \zeta & \text{sinon .} \end{cases}$$

où  $\zeta \in [0, 1]$ .

Il nous faut maintenant estimer  $s_0$  et  $\zeta$ .

### Algorithme EM 2

*Remarque : Nous nous contenterons ici d'indiquer les changements dans l'algorithme EM1. La méthode utilisée pour le produit de convolution ainsi que pour les données censurées reste valable ici.*

1. Tape E :

Si  $s_k > s_0$

$$\hat{t}_k(r) = \frac{\pi(s_k) f_{N_p}(n_k; s_k, \theta_p)}{\sum_{l=(1,2)} \pi(s_k) \Phi_l(n_k; s_k; \theta)}$$

Si  $s_k < s_0$

$$\hat{t}_k(r) = 0.$$

2. Tape M :

– La phase de maximisation pour les paramètres de propagation et d'amorçage est identique à la précédente.

Ce qui donne pour  $\zeta$ , le paramètre de  $\pi$  :

$$\max_{\zeta} \sum_{k=1 \dots N} 1_{s_k > s_0} (\hat{t}_k \log(\pi(s_k; \zeta)) + (1 - \hat{t}_k) \log(1 - \pi(s_k; \zeta)))$$

Il s'agit de la même maximisation que précédemment à la différence que seules les données où  $s_k > s_0$  sont utilisées.

### Recherche du meilleur seuil

Le modèle a ici été donné pour tout seuil  $s_0$ . Toutefois, celui-ci étant inconnu, il convient d'en donner une estimation. Pour ce faire, nous allons construire une grille sur l'ensemble des contraintes, puis celles-ci seront prises une à une et testées. Nous garderons le seuil donnant le maximum de vraisemblance le plus élevé.

Nous cherchons donc le seuil  $\hat{s}_0$  tel que :

$$\hat{s}_0 \in \underset{s_0 \in \text{Grille}}{\operatorname{argmax}} \prod_i f_N(n_i)$$

avec  $f_N$  définie par l'équation (4.11).

L'algorithme utilisé dans le cadre de cette recherche du meilleur seuil sera le suivant :

1. Découpage en grille de  $K$  seuils ( $K < N$ ), qui correspondent au niveaux de déformations appliqués à l'éprouvette.

*Remarque : le fait que le nombre de seuils à tester soit inférieur à  $N$  tient au fait que deux éprouvettes peuvent être testées aux mêmes niveaux de déformation.*

2. Pour chaque seuil, calcul de la vraisemblance du modèle : on obtient donc  $K$  vraisemblances associés aux différents seuils.
3. On retient le modèle avec la meilleure vraisemblance puisque tous les seuils possèdent la même modélisation et le même nombre de paramètres.

Le principal défaut de cette méthode est sa lenteur puisqu'elle nécessite de recalculer tous les estimateurs pour chaque seuil.

### Résultats sur simulation

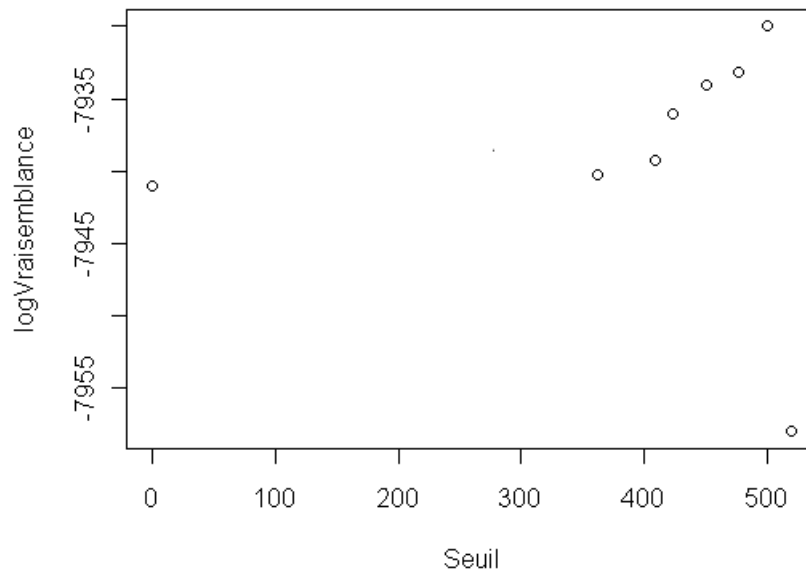


FIGURE 4.2 – Evolution de la logvraisemblance en fonction du seuil choisi.

La Figure 4.2 représente l'évolution de la vraisemblance pour différents seuils.



Tout d'abord on constate une augmentation de la vraisemblance jusqu'à atteindre un maximum, puis une chute brutale de la vraisemblance. Cette chute brusque correspond à l'apparition de points en « propagation pure » et donc l'apparition du phénomène de bi-modalité dans nos données. Ainsi le seuil fixé lors de la simulation est bien retrouvé par cette méthode.

La Figure 4.3 donne le quantile à 0,1% obtenu sur ces mêmes données grâce à notre modèle.

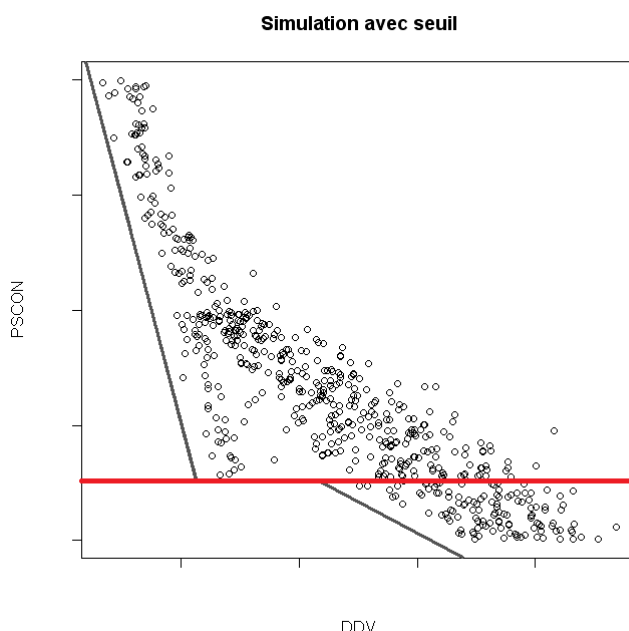


FIGURE 4.3 – Quantile à 0,1% obtenu sur données simulées.

Ce modèle a toutefois un gros désavantage. Le seuil est placé au niveau du dernier point classé en « propagation pure ». C'est-à-dire que, quel que soit le seuil réel, son estimateur sera placé au niveau de la valeur la plus basse obtenue dans les données. Or nous sommes dans un contexte de fiabilité et il apparaît qu'avoir un tel seuil fixé sur la valeur la plus basse ne permet pas d'obtenir un quantile à 0,1% consistant puisque l'estimation de ce seuil variera avec chaque jeu de données et le quantile à 0,1% du modèle changera donc fortement à chaque fois.

Ce modèle ne peut donc pas être retenu par la suite puisqu'il n'offre pas de garanties suffisantes quand au seuil à partir duquel  $\pi(S) = 0$ .

### 4.2.2 Proportion logistique

Ici, contrairement au  $\pi$  constant par morceau, l'hypothèse sera que plus la contrainte est faible plus le nombre d'éprouvettes amorçant au premier cycle est réduit. La nocivité des défauts est ici considérée comme différente pour chaque défaut et certains pourront propager sans amorçage initial à des contraintes plus faibles que d'autres.

Nous écrivons la dépendance de  $\pi$  en  $S$  grâce à un modèle logistique :

$$\pi(S) = \frac{e^{\alpha+\beta S}}{1+e^{\alpha+\beta S}}.$$

*Remarque : Le modèle avec un  $\pi$  logistique reste identifiable (voir annexe A).*

#### Algorithme EM 3

*Remarque : Nous nous contenterons ici d'indiquer les changements dans l'algorithme. La méthode utilisée pour le produit de convolution ainsi que pour les données censurées reste valable ici.*

1. Etape E :

$$\hat{t}_k(r) = \frac{\pi(s_k; \alpha, \beta) f_{N_p}(n_k; s_k, \theta_p)}{\sum_{l=(1,2)} \pi(s_k; \alpha, \beta) \Phi_l(n_k; s_k; \theta)}.$$

2. La phase de maximisation pour les paramètres de propagation et d'amorçage est identique aux précédentes.

Pour la maximisation de  $\alpha, \beta$  les paramètres de la logistique :

$$\max_{\alpha, \beta} \sum_{i=1 \dots N} \hat{t}_k \log(\pi(s_k; \alpha, \beta)) + (1 - \hat{t}_k) \log(1 - \pi(s_k; \alpha, \beta)).$$

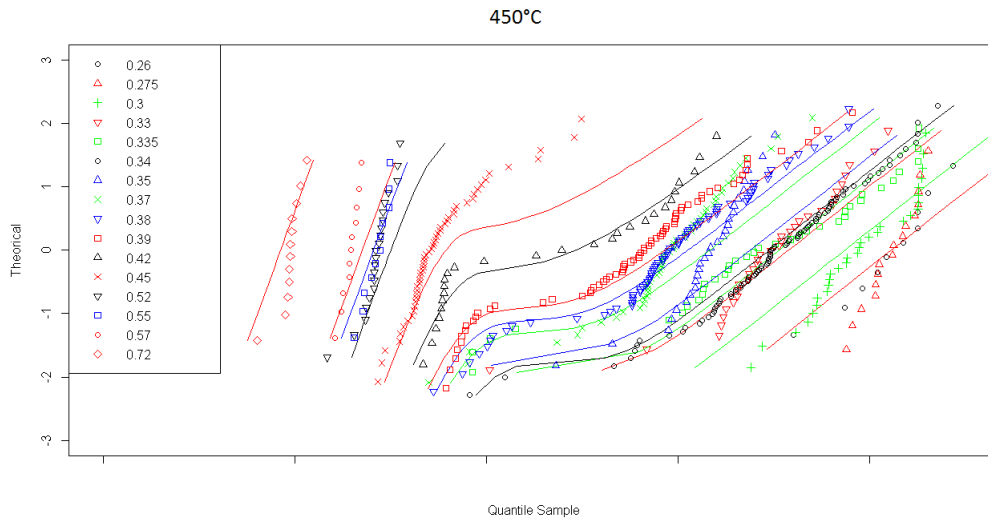
Ce qui, après simplification, donne :

$$\max_{\alpha, \beta} \sum_{i=1 \dots N} \hat{t}_k (\alpha + \beta s_k) + \log(1 + e^{\alpha + \beta s_k}).$$

#### Résultats sur données réelles

La Figure 4.4 représente le « quantile plot » des données. Les lignes continues représentent les prédictions de notre modèle pour chaque niveau de déformation. Ce graphique permet donc de vérifier visuellement l'adéquation des données à notre modèle.

Plusieurs constats peuvent être faits :

FIGURE 4.4 – quantile plot avec  $\pi(S)$  logistique.

- Pour les fortes déformations, l'adéquation est bonne, la partie propagation du modèle est bien retrouvée.
- La détection du passage du mode 1 ( $N = N_i$ ) au mode 2 ( $N = N_i + N_p$ ), visualisable par un saut est globalement bien retrouvée.
- Pour la partie amorçage et propagation, visualisable pour les faibles déformations, on constate des écarts entre la distribution des données et ce qui est modélisé. Cela s'explique par les nombreuses causes possibles d'amorçages qui mènent à d'importantes variations dans les données. Ce problème sera traité en fin de chapitre et des propositions d'amélioration seront données au chapitre 5.

La modélisation avec une proportion logistique est donc celle qui se rapproche le plus des données. (Cf. Table 4.2).

paramètre	$\alpha$	$\beta$	$a_i$	$b_i$	$\sigma_i$	$a_p$	$b_p$	$\sigma_p$
valeur obtenue	40	-20	20	-25	1	10	-5	0.2

TABLE 4.2 – Paramètres obtenus, ordre de grandeur

### 4.3 Calibrage de l'algorithme

Cette partie a pour but de déterminer quelles sont les bonnes conditions d'utilisation des algorithmes présentés précédemment. Pour cela nous

études le comportement de l'algorithme suivant différentes conditions sur les paramètres à notre disposition.

*Remarque : Cette étude sera réalisée pour le modèle avec une proportion logistique, les autres modèles, n'étant pas retenus, ne seront pas étudiés ici.*

### 4.3.1 Initialisation

La maximisation réalisée par l'algorithme EM garantit la convergence vers un maximum local. Ainsi une initialisation trop éloignée des vrais paramètres pourra amener à un « vidage de la classe amorçage-propagation » de notre modèle. C'est-à-dire que ne subsistera du modèle qu'une unique droite de propagation (voir Figure 4.5).

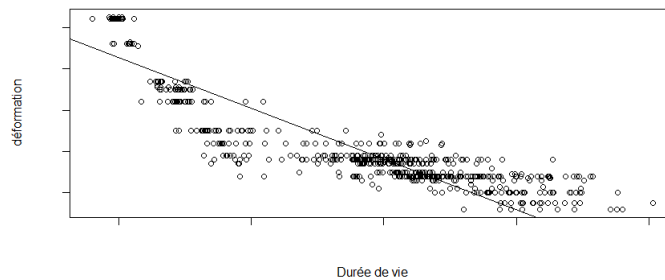


FIGURE 4.5 – Algorithme ayant vidé complètement le mode 2

#### small-em sur modèle

Plusieurs stratégies ont été proposées afin de contrer ce phénomène dans le cas de modèles de mélanges classiques, on citera par exemple les « small-em » (cf. Biernacki et al. [7]). Il s'agit de prendre plusieurs initialisations aléatoires et de leur appliquer quelques itérations de l'algorithme EM pour ne retenir que celle dont la vraisemblance est maximale à l'issue de ces itérations.

Ici nous découperons nos données aléatoirement en deux classes et initialiserons l'algorithme avec la moyenne de leurs coordonnées comme paramètre de position, une pente nulle et une proportion de 0,5. L'algorithme EM est ensuite itéré un certain nombre de fois, et toute cette procédure est répétée plusieurs fois. On garde enfin les paramètres qui donnent la meilleure vraisemblance parmi tous les jeux obtenus. Cette procédure a comme désavantage dans notre cas de nécessiter un grand nombre d'appels à l'algorithme EM.

Nous proposons donc, afin d'améliorer la rapidité d'obtention de cette initialisation, de reproduire cette stratégie sur un modèle simplifié de régression par classe. Ce modèle est décrit ci-dessous.

### small-em en régression par classe

La régression par classe fait l'hypothèse d'une partition des données en  $K$  classes, chaque classe étant composée de données  $Y$  en régression avec les données  $X$  et de son propre lot de paramètres  $\theta_K$ . Les classes étant inconnues, nous sommes en présence d'un algorithme avec données manquantes. Le modèle peut s'écrire :

$$Y \sim \sum_K \pi_K \mathcal{N}(a_K X + b_K, \sigma_K).$$

Nous pouvons donc ici aussi utiliser un algorithme de type EM afin de déterminer les paramètres et l'appliquer sur notre jeu de données, en prenant  $K = 2$ . Se faisant nous nous trouvons encore une fois devant un problème d'initialisation. Cependant la méthode de régression par classe est beaucoup plus rapide (un rapport de l'ordre de  $10^4$  entre les deux algorithmes). Ainsi appliquer une méthode « small-em » dans un premier temps sur l'algorithme de régression par classe pour pouvoir obtenir ensuite l'initialisation de notre algorithme plus lent est une bonne stratégie. Comme nous pouvons le voir dans le 4.3, les estimateurs issus de la régression par classe sont peu éloignés des estimateurs finaux en ce qui concerne les paramètres de propagation et d'amorçage. Pour le paramètre de proportion, nous recommandons de prendre  $\pi = 0,5$  quel que soit  $S$ .

paramètre	Ecart relatif
$a_i$	14%
$b_i$	20%
$\sigma_i$	5%

TABLE 4.3 – Comparaison des résultats de l'algorithme à ceux d'une régression par classe, écart relatif.

*Remarque : Les résultats sont semblables pour les données de propagation.*

### 4.3.2 Nombre de simulations Monte-Carlo

La détermination du nombre de points à utiliser pour le calcul de l'intégrale (3.3) est importante. Il y a un compromis à faire entre un faible nombre de points et donc une mauvaise précision dans l'approximation et un nombre élevé de points qui mène à un algorithme plus long. Dans notre cas, il apparaît qu'un nombre trop peu élevé d'itérations Monte-Carlo amène régulièrement des problèmes numériques, (l'intégrale pouvant alors être évaluée très proche de 0, provoquant une erreur au moment du passage au log.) Il semble que prendre un nombre de 10000 itérations soit toutefois suffisant pour contrer ce phénomène.

### 4.3.3 Nombre d'itérations de l'algorithme S-EM

La principale contrainte quand au nombre total d'itération de l'algorithme S-EM est de s'assurer de la convergence des estimateurs. Le graphique (4.6) nous offre la visualisation de l'évolution de la vraisemblance en fonction du nombre de boucles effectuées par l'algorithme EM.

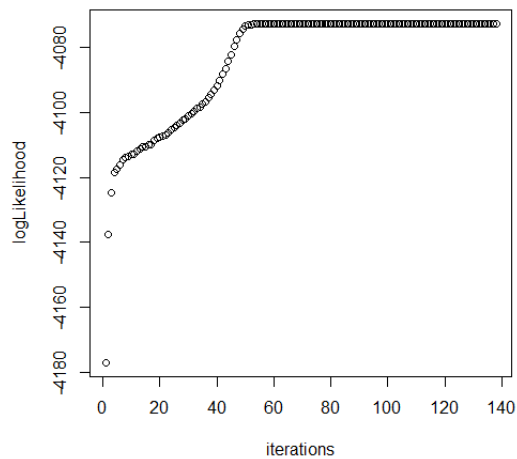


FIGURE 4.6 – Evolution de la logVraisemblance à chaque pas de l'algorithme

On s'aperçoit qu'après une rapide augmentation de la vraisemblance, l'algorithme semble se stabiliser à partir de la soixantième itération. Toutefois, un zoom (voir Figure 4.7) sur l'évolution de cette vraisemblance à partir de cette étape montre que la convergence n'est pas entièrement réalisée puisqu'on observe des fluctuations.

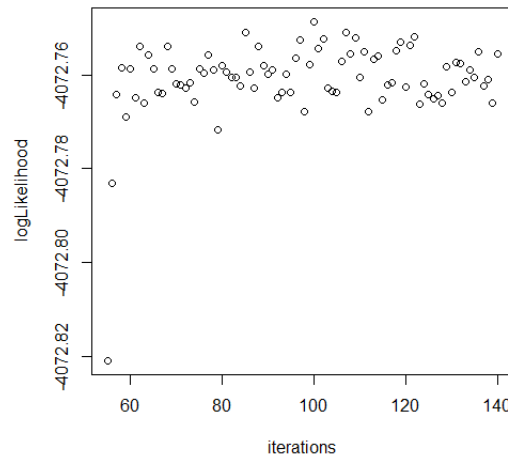


FIGURE 4.7 – Zoom sur l'évolution de la logVraisemblance

Celles-ci ont deux causes :

- l'utilisation d'un algorithme de Monte-Carlo pour l'estimation de la densité du deuxième mode introduit une dispersion dans l'estimation des paramètres que l'on retrouve dans ce graphique ;
- l'utilisation d'une étape Stochastique pour l'estimation des données censurées donne une convergence de l'algorithme non pas vers une valeur exacte mais vers les valeurs d'une chaîne de Markov dont les points stationnaires sont les paramètres voulus.

Afin d'obtenir notre estimateur, il nous faut donc considérer non pas la dernière réalisation de l'algorithme, mais une moyenne sur les dernières valeurs fournies.

Nous recommandons donc de prendre une « période de chauffe » d'environ 100 itérations pour atteindre la phase stationnaire, puis une cinquantaine d'itérations dont on fera la moyenne pour obtenir les différents paramètres.

#### 4.3.4 Nombre de points disponibles

Le but ici est de déterminer le nombre minimal de données pour lequel l'algorithme présente un bon comportement. Il s'agit du seul paramètre étudié qui ne soit pas un paramètre de contrôle de l'algorithme. Seuls les résultats pour la proportion  $\pi$  et le paramètre d'ordonnée à l'origine de l'amorçage  $\theta_a$  sont visualisables ici. Les autres paramètres sont visualisables en annexe (C).

Le tableau (4.4) donne le nombre de fois où l'algorithme a totalement

vidé la classe de propagation pure. Il faut toutefois noter que même dans ces conditions, les paramètres de propagation et d’amorçage restent estimables puisqu’ils apparaissent conjointement dans le mode 2 de notre mélange.

nombre de points	pourcentage de classes vides
20	30%
50	12%
100	0%
200	0%
400	0%

TABLE 4.4 – Sur 100 simulations, risque de dégénérescence (la classe propagation pure a été vidée)

nombre de points	écart-type
20	1.27
50	0.8
100	0.57
200	0.56
400	0.62

TABLE 4.5 – Sur 100 simulations, comportement de l’estimateur du paramètre  $a_p$  obtenu, le coefficient à l’origine de la propagation.

On notera qu’à moins de 50 points, l’algorithme a tendance à vider les classes. L’estimation des paramètres d’amorçage et de propagation semblent se stabiliser à partir des 50 points. Il apparaît donc qu’un minimum de 50 points doivent être demandés afin d’assurer une bonne estimation des paramètres dans le cadre d’un modèle de mélange amorçage-propagation.

*Remarque : Les résultats des simulations pour les autres paramètres sont disponibles en Annexe C*

## 4.4 Réduction du risque, quel(s) quantile(s) utiliser ?

On l’a vu, les trois modélisations proposées ( $\pi$  constant, par morceau ou logistique) mènent chacune à un quantile distinct. Cependant la question reste à savoir lequel semble le plus approprié dans un cadre de recherche



du « 0 défaut ». La modélisation avec  $\pi$  constant par morceau a de suite été écartée puisque n'apportant pas les garanties de fiabilité suffisante. Reste maintenant à faire le choix entre le modèle du  $\pi$  constant ou du  $\pi$  logistique.

Nous noterons tout d'abord que les estimateurs des paramètres d'amorçage et de propagation sont proches dans les deux modélisations. Ainsi la principale différence vient de la forme du  $\pi$ . Nous avons vu que, dans les deux cas, l'estimation de sa valeur est la moins stable de tous les paramètres. L'erreur d'estimation est donc importante sur cette quantité. Par ailleurs, si on se réfère aux travaux de Sakai (cf. [39]), les spécialistes des matériaux considèrent son estimation comme difficile et ne préconisent pas de forme particulière à appliquer. La modélisation est toutefois plus proche des données dans le cas d'une modélisation logistique de  $\pi$ .

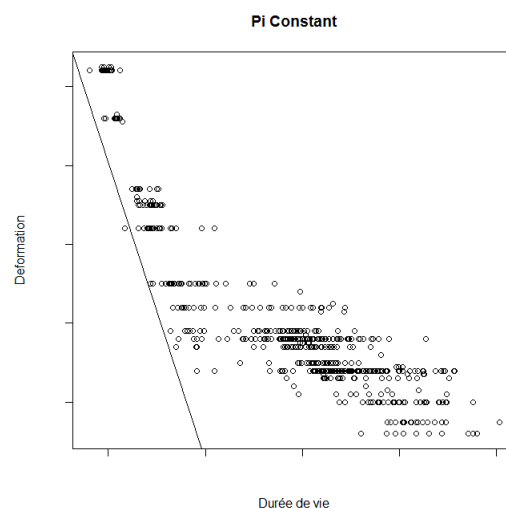


FIGURE 4.8 – Quantile à 0,1% avec  $\pi$  constant

La Figure 4.8 (respectivement 4.9), représente le quantile obtenu pour une modélisation à  $\pi$  constant (respectivement logistique). On constate que le quantile pour  $\pi$  constant est en fait très proche du quantile que l'on aurait obtenu en ne considérant que la « propagation pure ». Tandis que le quantile de la deuxième courbe suit bien plus les données. Toutefois le passage du mode de propagation au mode « amorçage + propagation » est difficile à déterminer au vu des incertitudes à la fois sur la forme du  $\pi$  et sur la difficulté de son estimation. C'est pour cela que Jha (cf. [25]) considère que le  $\pi$  constant est la meilleure réponse à apporter au problème. Puisque les raisons exactes d'un passage du mode 1 au mode 2 sont inconnues. Dans un cadre de fiabilité le plus sûr est donc de considérer uniquement le plus dégradé des

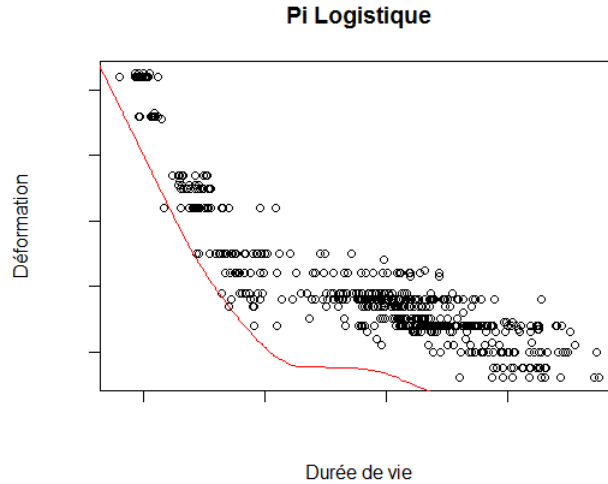


FIGURE 4.9 – Quantile à 0.1% avec  $\pi$  variable

modes pour en faire un quantile.

Si cette option à l'avantage de la sûreté, elle présente toutefois l'inconvénient d'être assez éloignée de la réalité du risque qu'encourt une éprouvette au cours des essais à faible déformation.

La solution pour résoudre ce problème serait donc d'acquérir une connaissance suffisante sur les raisons du passage du mode 1 au mode 2 d'un point de vue matériau. Ainsi pourrait se greffer à la modélisation réalisée un seuil externe déterminé par les ingénieurs matériaux afin de s'approcher le plus possible du risque réel de défaillance de l'éprouvette lors des essais aux faibles déformations (cf. Thieulot-Laure [46]).

## 4.5 Application à différentes températures

La température de 450°C n'est pas l'unique testée de la base de données. Il s'agit toutefois de celle où la bi-modalité est la plus visible dans nos données. Nous allons maintenant nous intéresser au comportement de notre algorithme pour ces autres températures.

Le tableau ci-dessous retrace l'évolution des paramètres d'amorçage et de propagation en fonction de la température. Si les coefficients de propagation semblent être similaires et peu variés, on constate que ceux d'amorçage ont des comportements fortement différents. Ainsi on peut former deux groupes dans les données : 450°C et 550°C ont un comportement similaire dans le sens où leurs coefficients  $b_i$  et  $\sigma_i$  sont proches. Cela signifie que les droites

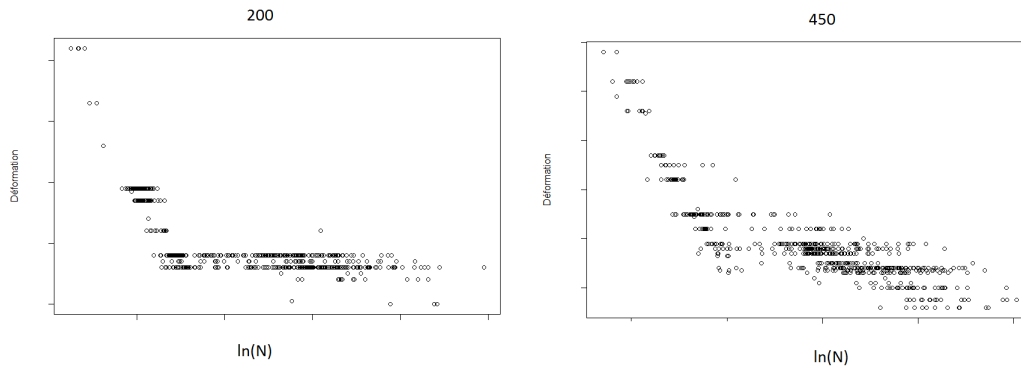


FIGURE 4.10 – Comparaison des données à 200 et 450°C

représentatives de leurs moyennes en amorçage sont parallèles sur un graphique. Tandis que les coefficients à 300°C et 200°C sont fortement éloignés avec une variance plus importante de ceux du premier groupe. On peut, en fait, observer directement ce phénomène sur les données (cf. Figure 4.10), où les données classées en « amorçage + propagation » sont bien plus étalées en durée de vie à 200°C qu'à 450°C (cf. Table 4.6)

température	$a_i$	$b_i$	$\sigma_i$	$a_p$	$b_p$	$\sigma_p$
550°C	0.7	0.4	0.01	-0.2	-0.1	0.02
450°C	0	0	0	0	0	0
350°C	5.2	-19	.82	0.5	-0.6	-0.03
200°C	3.2	-17	.43	1	-1.4	-0.03

TABLE 4.6 – Estimation des paramètres aux différentes températures. Ecart par rapport à 450°C

La Figure 4.11 présente les quantiles à 0.1% pour les différentes températures. On constate l'augmentation globale de la durée de vie en amorçage avec l'augmentation de la température. Toutefois le phénomène inverse semble se produire pour la propagation, le quantile est supérieur lorsque la température est faible. Cependant les variations des paramètres de propagation en fonction de la température sont faibles. Cela peut partiellement être recoupé par l'analyse réalisée par Fournier et Pineau [22], dont l'une des conclusions est que la température influe principalement sur la durée de vie en amorçage et peu sur la durée de propagation, bien que la variation se fasse ici en sens inverse de ce qui est observé dans leur article (une plus longue durée de vie à la plus basse température). Cependant les températures

testées ne sont pas identiques puisque nos essais vont de 200 à 550 °C là où ils étudient le comportement à 25°C et 550°C. Par ailleurs, au vu de l'étendue de la base de données Snecma, il est probable que la structure du matériau des éprouvettes testées dans leur article soit plus homogène.

Le croisement des quantiles est réalisé pour une déformation d'environ 0.3% ce qui est aussi la zone de transition entre « propagation pure » et « propagation + amorçage ».

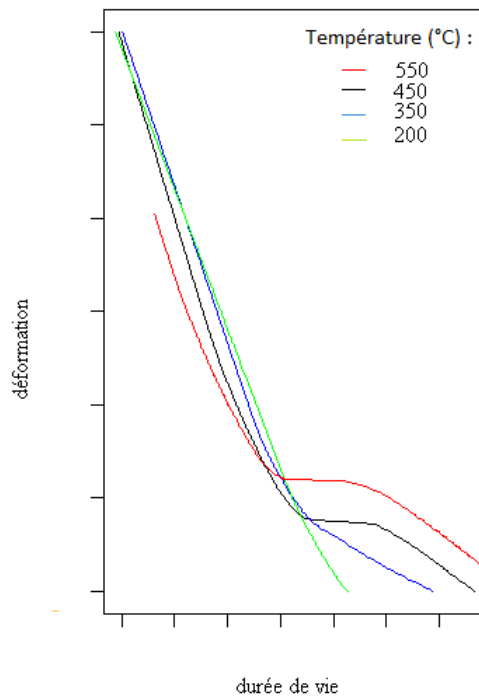


FIGURE 4.11 – Quantile à 0.1% pour les différentes températures

Si on constate des différences de comportement aux différentes températures, réaliser un modèle global multi-température est difficile. En effet ce graphique permet de rendre compte de la non linéarité du phénomène puisque les comportements en amorçage à 450°C et 550°C sont fortement différents de ce qui est observé à 350°C et 200°C. Ainsi une modélisation de la température doit elle passer par des modèles de mécanique des matériaux. Le simple ajout « naïf » d'une relation linéaire en fonction de la température ne suffisant pas à prendre en compte la diversité observée.

## 4.6 Conclusion

Dans ce chapitre nous avons donc mis en œuvre l'estimation des paramètres du modèle de mélange « amorçage-propagation ». Les résultats obtenus sur les données Snecma font apparaître que la probabilité  $\pi$  d'amorcer au premier cycle dépend de la contrainte  $S$ . Nous avons alors testé différents modèles pour  $\pi$ . Le modèle retenu étant le modèle logistique. Par ailleurs la procédure d'estimation a été calibrée à partir de données simulées. Le modèle de mélange « amorçage-propagation » permet finalement une classification des éprouvettes en fonction du type d'amorçage (rapide/long).

### 4.6.1 Prédiction

La première application du modèle (celle pour laquelle il a été créé) est donc la prédiction de durées de vie des éprouvette en fatigue. En estimant de manière fiable les paramètres d'amorçage et de propagation, il permet de retrouver aisément des quantiles à faible probabilité de rupture de l'éprouvette. On constate d'ailleurs qu'il n'y a pas de groupes de points qui ressortent du quantile à 0,1% (cf. Figure 4.9) comme ce pouvait être le cas des modèles statistiques présentés au §2.2.2. Toutefois ce modèle ne permet pas directement une prédiction de durée de vie de la pièce en vol puisqu'il y a unicité du volume des éprouvettes testées. Ainsi nous ne pouvons pas prédire ce qui se produit au niveau d'une pièce complète.

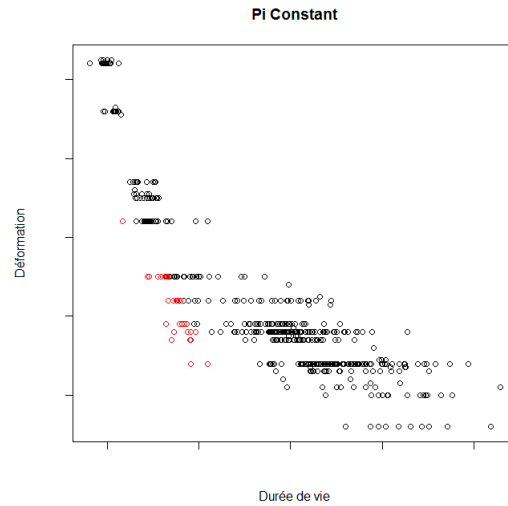
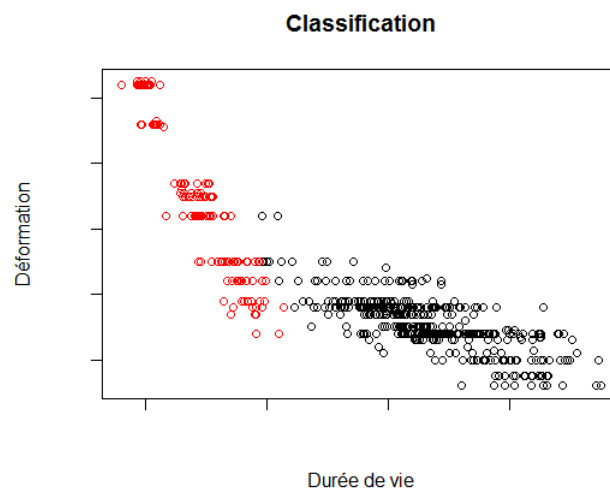
### 4.6.2 Classification des données

La classification est réalisée par la règle du maximum a posteriori (MAP) : Une fois l'estimation du vecteur des paramètres  $\hat{\theta}$  effectuée, on détermine la meilleure partition des observations en attribuant à chaque individu la classe pour laquelle il a la plus forte probabilité d'appartenance. Pour cela, les probabilités conditionnelles  $P(n_i \in P_k | n_1, \dots, n_N)$  que l'observation  $n_i$  appartienne à la classe  $k$  sachant l'ensemble des observations sont calculées. On a par le théorème de Bayes,

$$t_{ik}(\hat{\theta}) = \frac{\hat{\pi}(s_k; \hat{\alpha}, \hat{\beta}) f_{N_p}(n_k; s_k, \hat{\theta}_p)}{\sum_{l=(1,2)} \pi(s_k; \hat{\alpha}, \hat{\beta}) \Phi_l(n_k; s_k; \hat{\theta})}.$$

Chaque observation est finalement attribuée à la classe pour laquelle la probabilité conditionnelle est la plus grande :

$$z_{i1} = \begin{cases} 1 & \text{si } t_{i1}(\hat{\theta}) > 0.5; \\ 0 & \text{sinon .} \end{cases}$$

FIGURE 4.12 – Classification des données avec  $\pi$  constantFIGURE 4.13 – Classification des données avec  $\pi$  variable

On observe que la classification n'est pas la même dans les deux cas. Le modèle avec proportion constante ne classe en pratique qu'un petit nombre de points dans le groupe des amorçages en propagation pure. Or, lorsque la déformation est importante l'amorçage est faible et le groupe de données devrait donc être classé dans le mode 1. Le modèle logistique fournit donc une classification plus proche de ce qui peut être attendu par un expert matériau. A la décharge du modèle constant, il faut toutefois noter que les deux distri-

butions (celle de propagation pure et celle d'amorçage + propagation) sont très semblables pour ces fortes déformations. Les différences de classification se jouent donc sur la valeur de  $\pi$  dans ces niveaux. Or le  $\pi$  constant étant global, il ne peut pas agir spécifiquement sur les fortes déformations d'où les difficultés de classification des points à ces niveaux.

Une des hypothèses métallurgiques est que les points ayant amorcés au premier cycle le font en surface. Comme certains essais ont donné lieu à une fractographie qui a permis de déterminer le lieu de l'amorçage, nous pouvons donc comparer le résultat de notre classification avec ces données.

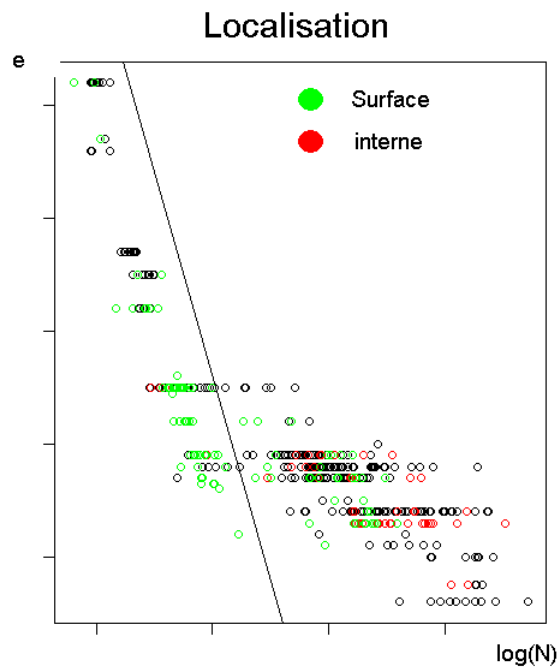


FIGURE 4.14 – Localisation des amorçages surfaciques et internes

Comme attendu, on constate que le mode d'amorçage interne est quasi-intégralement représenté dans le mode amorçage + propagation. Toutefois, les amorçages surfaciques, dont on pouvait attendre qu'ils soient présents uniquement dans le mode de propagation pure sont en fait situés dans les 2 modes. Ainsi, il semble que les raisons exactes d'un amorçage rapide soient plus complexes qu'une simple distinction entre amorçage en surface et amorçage interne.

### 4.6.3 Avantages et inconvénients de la modélisation

En premier lieu, la modélisation obtenue rend bien compte du phénomène de bi-modalité observable. Le modèle est « proche » de ce qui est visualisé et ce, sur toute la plage de déformations testées.

Par ailleurs, le modèle se base sur une approche mécanique de la rupture, la distinction entre amorçage rapide et amorçage lent. Ainsi, le modèle peut facilement être interprété par les ingénieurs.

Ces points positifs sont toutefois à nuancer par quelques manques. Tout d'abord, considérer qu'il n'existe qu'un seul type d'amorçage dans la partie « amorçage + propagation » semble insuffisant. Nous pouvons en effet constater sur la Figure 4.14 qu'on trouve à la fois des amorçages en surface et interne. Par ailleurs on a du mal, même en connaissant les raisons d'amorçage, à estimer les proportions (voir [39] et [25] par exemple). Cette difficulté est présente dans notre modèle, ce qui amène une incertitude quand au quantile à utiliser.

#### Voies d'amélioration

La première voie d'amélioration sera développée dans le chapitre 5. Il s'agit de prendre en compte la possibilité d'amorçages multiples dans nos données. Si on se réfère à la Figure (4.15). On constate en effet que les données pour une déformation de 0.36 intersectent celles de 0.38. Or cette intersection n'est pas prévue par notre modélisation. Nous faisons en effet l'hypothèse de variances égales et de moyennes décroissantes avec l'augmentation de la déformation. En d'autres termes les deux courbes devraient être parallèles sur un « quantile plot ».

Afin de pallier ce problème, nous proposons dans le chapitre suivant de rajouter des informations matériau supplémentaires. Celles-ci devant amener à considérer la base de données Snecma non pas comme un unique bloc de matériau Inconel homogène. Mais comme étant représentatif des différences que peut subir le matériau en fonction des variations dans le déroulement de sa fabrication (en particulier son forgeage).

---



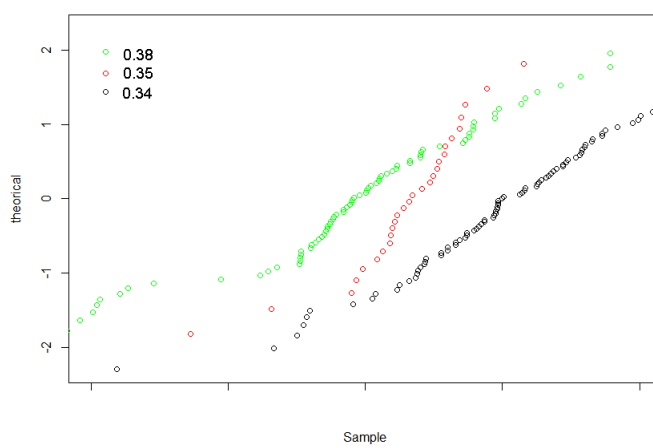


FIGURE 4.15 – QQplot des données pour 3 niveaux de déformation



## Chapitre 5

# Prise en compte d'informations de production

Notre modèle intègre le fait que la durée de vie est réduite en moyenne quand la déformation augmente. Or nous observons que ce principe n'est pas respecté sur certaines données. Ce phénomène suggère l'existence de comportements différents en terme de caractéristiques mécaniques. L'idée de ce chapitre est donc d'identifier des sous-populations à partir d'information supplémentaire : les données de production ainsi que des données sur les pièces elles-mêmes, comme le moteur (et donc la gamme de production) auquel elles sont destinées par exemple, et nous tenterons de les intégrer dans notre modèle.

### 5.1 Forgeage d'une pièce de turboréacteur

Le principe du forgeage n'a que peu varié depuis l'antiquité. Il s'agit de réaliser une déformation du métal, à chaud ou à froid, par l'utilisation d'un outil de choc. Les presses hydrauliques faisant désormais office de marteaux et d'enclumes. Toutefois le procédé en lui-même s'est complexifié. Les exigences de sécurité actuelles imposent à la fois une plus grande précision d'action lors des différentes étapes de fabrication, mais aussi un contrôle qualité tout au long du processus. Le processus de forgeage d'un disque de turbine est décrit dans le graphique (5.1).

Il consiste en plusieurs déformations pour mener une billette initiale à une matrice prête à être usinée afin d'obtenir la pièce finale. Cette billette initiale est d'abord écrasée afin d'obtenir la bonne épaisseur. Le débouchage puis le laminage permettent ensuite d'amener la pièce à la bonne circonférence. Une dernière opération de matriçage permet au matériau de prendre une forme

---

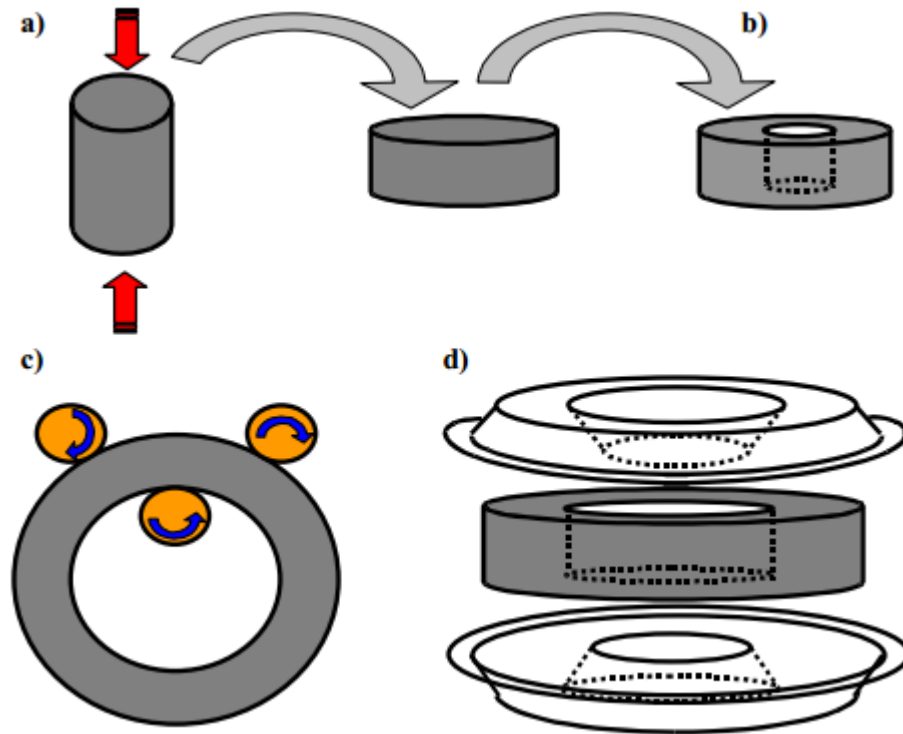


FIGURE 5.1 – Principe de forgeage d'une pièce de turboréacteur

brute plus proche de la pièce finale.

L'étape de forgeage est importante puisque c'est d'elle que dépendent les propriétés du matériau telle que la taille de grain ainsi que les différentes contraintes résiduelles présentes au sein du matériau.

Les paramètres influençant la taille de grains finale de la pièce sont nombreux. On pourra par exemple citer la vitesse de refroidissement, le niveau de déformation ou le niveau de contrainte appliqué lors du forgeage. Toutefois ces données ne sont pas directement accessibles en production. Ce dont nous disposons, ce sont des informations directement manipulables par le technicien en charge de la fabrication, mais pas de données locales au sein de la pièce. Il s'agira par exemple du temps d'attente entre l'étape a) et l'étape b) du graphique 5.1.

## 5.2 données disponibles

1. Les informations relatives à la pièce :

---

Plusieurs types d'informations sont présentes pour le matériau dont est issue l'éprouvette. Toutefois la base n'est que partiellement remplie en ce qui concerne les données de production et les informations sur la microstructure de l'éprouvette.

- Moteur (A,B,C,D)
  - Type de pièce (Disques de turbine, Disques de compresseur, Disque Laby, ...)
  - Lot matière
  - Fournisseur de matière
  - Forgeron sous-traitant
2. Les informations relatives à l'essai :
- Labo de test
  - Partiel : taille de grain
  - Partiel 40% : type d'amorçage (carbure, nitrure, clivage de grain...)
  - Partiel : zone d'amorçage (interne, surface, sous-couche ...)
3. Les informations relatives à la production :
- Durée de forgeage
  - Durées d'attente
  - Différentes températures (outils, surface du matériau, ...)
  - Matricule des opérateurs
  - Effort appliqué
  - Hauteur finale de la pièce

*Remarque : Les informations notées partiel ne sont pas présentes pour toutes les éprouvettes.*

### 5.2.1 Informations de production

Les données de production ont ceci de particulier qu'elles ne sont disponibles que pour la pièce complète, contrairement aux données d'essai qui sont des informations relatives à l'éprouvette. Ainsi, elles ne sont que peu représentatives de ce qui se passe au niveau local sur la structure de l'éprouvette. Afin d'obtenir une modélisation plus précise, une analyse partielle (sur un nombre restreint de données) a été réalisée. Elle consiste à reprendre les données de production, et à simuler tout le processus à l'aide du logiciel FORGE qui traite le problème par éléments finis. Cela permet, à partir d'observations globales sur la pièce d'obtenir les valeurs locales théoriques en ce qui concerne différents paramètres physiques tels que :

- La température locale maximum
  - La déformation locale
  - La vitesse de refroidissement
-

— . . .

À partir de là une Analyse en Composantes Principales (ACP) permet une visualisation des données en deux dimensions. La projection des données sur les axes principaux est visualisable sur la Figure 5.2 (Pour plus d'information sur les axes, on pourra consulter [21])

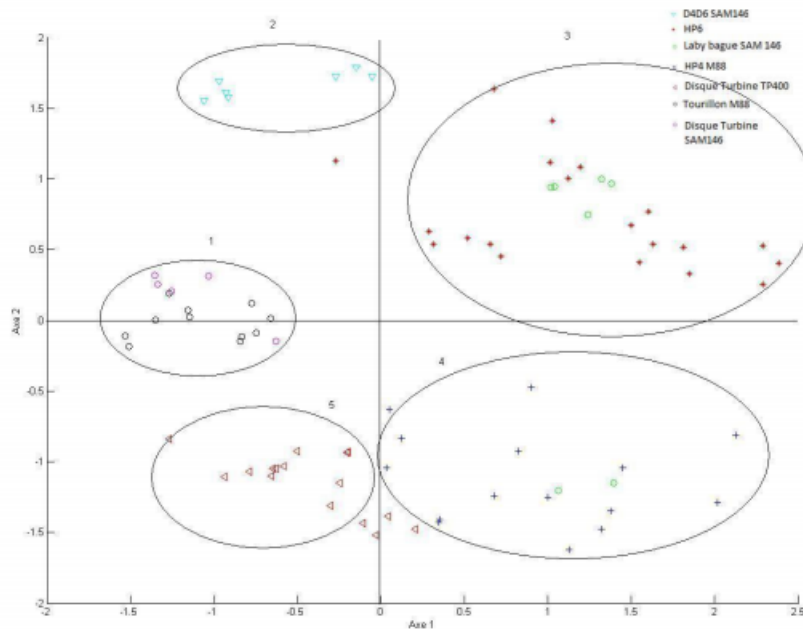


FIGURE 5.2 – Projection des types de pièce sur le plan principal

On constate qu'on peut facilement, à l'aide de cet ACP reconstituer les groupes de pièces de la base. L'information sur les données de production, même prise de manière locale, est en fait en grande partie contenue dans la gamme de forgeage, c'est-à-dire la pièce. Par la suite, nous nous restreindrons donc à l'étude des différentes pièces au lieu de l'étude complète des informations de production. Le principal avantage de prendre en entrée les pièces et non les informations locale de production est que ces données sont faciles d'accès et la modélisation proposée pourra donc être réutilisée dans le cadre d'un suivi régulier des points de fatigue, alors que les aux données locales de production demandent un post-traitement assez lourd.

---

## 5.3 Modèle Hiérarchique

Le croisement des courbes laisse penser qu'au moins deux comportements différents non identifiés sont présents dans la base de donnée. Ceux-ci étant inconnus, nous devons donc introduire un nouveau niveau de données manquantes. La modélisation se fera à l'aide d'un modèle hiérarchique ou autrement dit, d'un modèle de mélange de mélange. La première couche de mélange sera la classification du type d'éprouvette, nommés arbitrairement « verts » ou « bleus ». La deuxième couche, une fois que le type d'éprouvette est déterminé étant la classification entre éprouvette à amorçage rapide et éprouvette à amorçage lent. Cette modélisation par modèle de mélange de mélange existe déjà dans le cadre de composants de régressions. Ils sont généralement nommés mélanges d'experts et sont présentés ci-dessous.

### 5.3.1 Modèle de mélange d'experts

Au même titre que CART (cf. Breiman [11]), le modèle de mélange hiérarchique d'experts (cf. Jordan et Jacobs [28]) tente de faire une classification de données (ici en régression linéaire). Le but est donc de déterminer les critères qui permettent de classifier les données en différents jeux comparables. À la différence de CART qui impose des frontières fixes (une fois le seuil dépassé on change de groupe), le modèle de mélange d'expert en régression permet d'avoir des frontières souples entre les différentes classes de données, c'est-à-dire que l'on ne passe pas directement de la classe 1 à la classe 2 mais progressivement, la proportion de données dans la classe 1 diminuant au détriment de la classe 2. L'exemple donné par le graphique 5.3 permet de mieux comprendre l'intérêt de ce type de modèle.

### 5.3.2 Modèle et notations

Au lieu de prendre comme base un modèle linéaire, nous reprenons donc l'idée du mélange d'expert avec comme maillon initial notre mélange « amorçage propagation ». Le but principal étant de déterminer quelles variables agissent sur la classification des données « vertes » ou « bleues ». L'utilisation de données qualitatives nécessite un post-traitement afin de les rendre utilisables. Celui-ci est donné ci-dessous.

#### Données qualitatives

Nos éprouvettes sont décrites à l'aide de  $p$  variables qualitatives. Les données brutes se présentent donc sous la forme d'un tableau à  $N$  lignes et

---

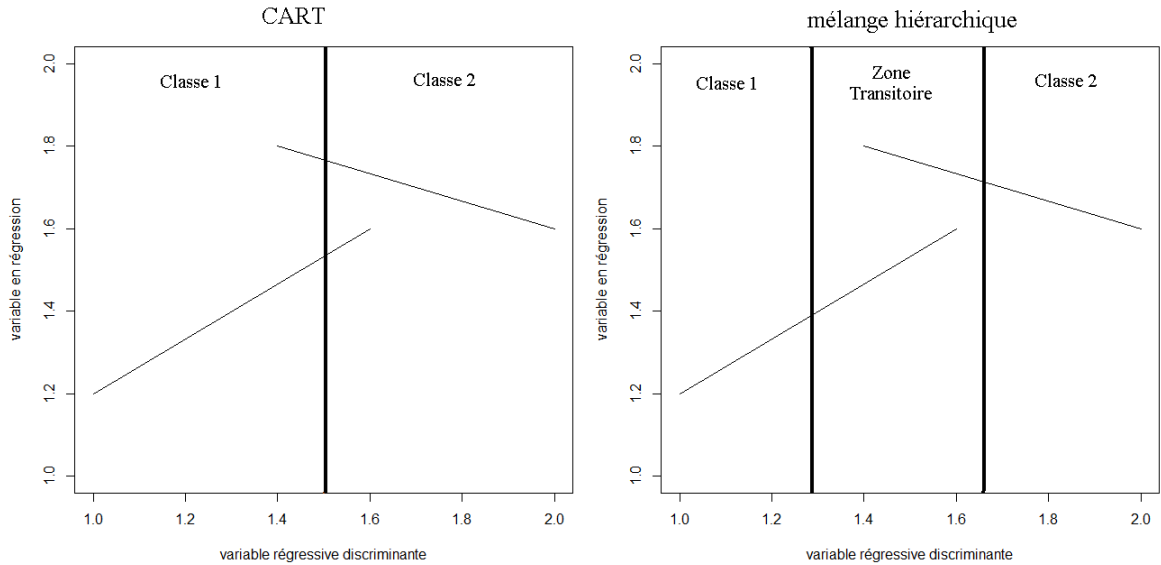


FIGURE 5.3 – Différence entre CART et les modèles de mélange hiérarchique

$p$  colonnes. Les éléments de ce tableau sont des codes sur lesquels aucune opération arithmétique n'est autorisée. La forme mathématique utile sur de tels tableaux est le tableau disjonctif des indicatrices des  $p$  variables obtenu en juxtaposant les  $p$  tableaux d'indicatrices de chaque variable.

Si nous prenons par exemple le tableau composé des quatre éprouvettes suivantes, décrites par leur moteur, le forgeron et le groupe de taille de grain. :

$$\begin{bmatrix} A & snecma & gros \\ A & autre & gros \\ B & snecma & gros \\ C & autre & petits \end{bmatrix}.$$

Il correspond donc à quatre observations de trois variables pouvant prendre respectivement 3,2,2 modalités. Ce tableau engendre le tableau disjonctif suivant à quatre lignes et sept colonnes :

$$\left[ \begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 \end{array} \right]$$

La somme des éléments de chaque ligne est égale à  $p$ , nombre de variables. La somme des éléments d'une colonne donne l'effectif marginal de la modalité correspondante. Le tableau disjonctif complet de notre jeu de données sera noté  $Q$ . Il a pour principal avantage de pouvoir être utilisé dans nos formules.



### Modèle

La durée de vie d'une éprouvette peut s'écrire de la manière suivante :

$$N = g_v * N_v + g_b * N_b. \quad (5.1)$$

L'équation (5.1) décrit le fait qu'une éprouvette peut appartenir, soit aux éprouvettes de type « vert », soit aux éprouvettes de type « bleu ». (Les noms verts et bleus n'ont pas de signification mais sont présents pour une meilleure visualisation). Ainsi l'éprouvette aura une certaine probabilité  $g_v$  d'appartenir au groupe des éprouvettes « vertes », groupe ayant pour durée de vie  $N_v$  avec le modèle d'amorçage-propagation présenté au chapitre précédent.

On a donc :

$$N_v = g_{1|v} * N_p + g_{2|v} * (N_p + N_{av})$$

$$N_b = g_{1|b} * N_p + g_{2|b} * (N_p + N_{ap})$$

De plus, pour chaque éprouvette  $k$ , la probabilité d'appartenir aux verts ou aux bleus sera modélisée par une distribution logistique fonction des données qualitatives :

$$\forall k \in 1 \dots N$$

$$g_{vk} = \frac{e^{a_v + (b_v)^{T * Q_k}}}{1 + e^{a_v + (b_v)^{T * Q_k}}}$$

$$g_{bk} = \frac{1}{1 + e^{a_v + (b_v)^{T * Q_k}}}$$

avec  $Q_k$  :  $k$ -ième ligne du tableau disjonctif complet.

Connaissant le groupe, la probabilité d'avoir un amorçage rapide sera modélisée par :

$$g_{1k|v} = \frac{e^{a_{1|v} + (b_{1|v}) * S_k}}{1 + e^{a_{1|v} + (b_{1|v}) * S_k}}$$

$$g_{2k|v} = \frac{1}{1 + e^{a_{1|v} + (b_{1|v}) * S_k}},$$

qui ne dépend que de la contrainte  $S_k$ .

### 5.3.3 Schéma

Le modèle hiérarchique est facilement visualisable sous forme d'arbre (voir graphique 5.4). L'arbre issu de notre modélisation est donc le suivant. Les paramètres d'amorçage sont obtenus par bloc (celui des « verts » ou des « bleus »).

Le but est donc bien de retrouver, sans connaissance a priori, deux blocs de données différents auxquels on applique notre modèle. La différenciation se faisant grâce aux données qualitatives de production introduite dans la

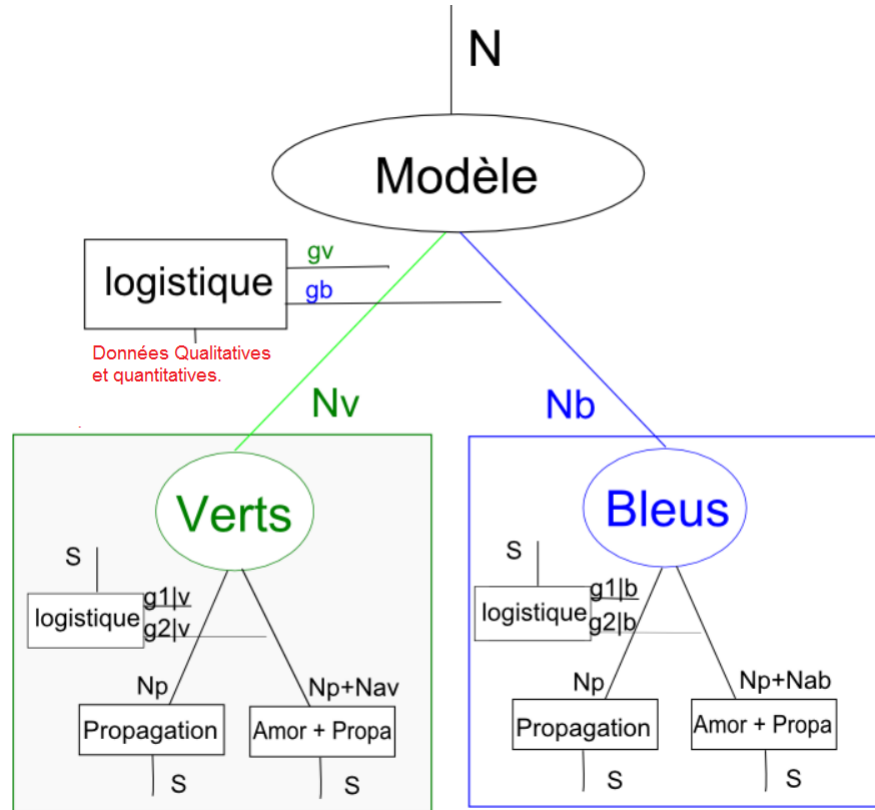


FIGURE 5.4 – Schéma du modèle hiérarchique

logistique qui donne une probabilité d'appartenance aux « verts » ou aux « bleus ».

## 5.4 Algorithme EM

Nous sommes donc en présence de deux types de données latentes. Tout d'abord, la classe à laquelle appartient l'éprouvette, classe nommée pour l'instant « vertes » ou « bleues » puisque nous ne possédons pas d'informations dessus a priori. Ensuite au sein de chaque classe, il y a un découpage entre les éprouvettes ayant amorcé au premier cycle et les autres. L'étape E nécessite donc le calcul d'au moins deux probabilités :

- Tout d'abord la probabilité d'être dans le groupe « vert » ou « bleu » d'éprouvettes.
- La probabilité connaissant le groupe de l'éprouvette d'avoir amorcé ou non rapidement.

La phase de maximisation reprend celle d'un modèle de mélange et consiste à maximiser la vraisemblance complétée à l'aide des deux probabilités calculées à l'étape E.

*Remarque : Afin d'améliorer la vitesse et la stabilité de l'algorithme, les paramètres de propagation seront ceux obtenus au chapitre précédent.*

### 5.4.1 Étape E

Calcul des probabilités conditionnelles :

- $i$  servira d'indexation pour les verts ou les bleus :  
 $i \in \{v, b\}$
- $j$  aura le rôle d'indexateur pour le groupe propagation (1) ou propagation + amorçage (2).

Nous calculons ainsi différentes probabilités :

- la probabilité d'être dans le groupe bleu ou vert sachant les données et une valeur courante des paramètres :

$$t_{ik}^{(r)} = \frac{g_i(Q_k) \sum_j g_{j|i}(S_k) \phi_{ij}(N_k | \theta_{ij}^{(r)}, S_k)}{\sum_i g_i(Q_k) \sum_j g_{j|i}(S_k) \phi_{ij}(N_k | \theta_{ij}^{(r)}, S_k)};$$

- la probabilité d'être dans le groupe propagation ou propagation + amorçage sachant l'appartenance aux verts ou aux bleus :

$$t_{jk|i}^{(r)} = \frac{g_{j|i}(S_k) \phi_{ij}(N_k | \theta_{ij}^{(r)}, S_k)}{\sum_j g_{j|i}(S_k) \phi_{ij}(N_k | \theta_{ij}^{(r)}, S_k)};$$

- la probabilité jointe, qui découle des deux probabilités ci-dessus :

$$t_{ijk}^{(r)} = \frac{g_i(Q_k) g_{j|i}(S_k) \phi_{ij}(N_k | \theta_{ij}^{(r)}, S_k)}{\sum_j g_i(Q_k) g_{j|i}(S_k) \phi_{ij}(N_k | \theta_{ij}^{(r)}, S_k)};$$

### 5.4.2 Étape M

Maximisation de la vraisemblance complétée, qui s'écrit :

$$l_c = \sum_k \sum_i \sum_j t_{ijk} \{ \log(g_i(Q_k)) + \log(g_{j|i}(S_k)) + \log(\phi_{ij}(N_k | \theta_{ij}, S_k)) \}.$$

Ce qui nous conduit aux trois maximisations suivantes :

- $\theta_{ij}^{(r+1)} = \underset{\theta_{ij}}{\operatorname{argmax}} (\sum_k t_{ijk} \log(\phi_{ij}(N_k | \theta_{ij}, S_k)))$ ;
- $\eta^{(r+1)} = \underset{\eta}{\operatorname{argmax}} (\sum_k \sum_i t_{ik} \log(g_{ik}))$ ;
- $\eta_{\cdot|i}^{(r+1)} = \underset{\eta_{\cdot|i}}{\operatorname{argmax}} (\sum_k \sum_j t_{jk|i} \log(g_{jk|i}))$ .

### 5.4.3 Résultats

Nous discuterons ici les résultats en appliquant notre algorithme de recherche des estimateurs des paramètres du modèle hiérarchique à la fois aux données simulées et aux données réelles.

#### Données Simulées

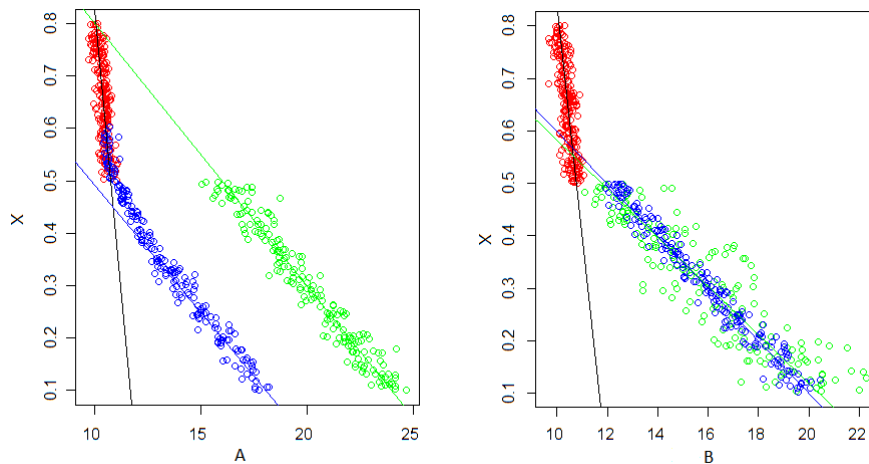


FIGURE 5.5 – Simulation de deux jeux de données suivant un modèle hiérarchique. A) position éloignée, B) position rapprochée

Nous avons simulé deux jeux de données représentés Figure 5.5. Un premier jeu de données où les positions des groupes « bleus » et « verts » sont assez éloignées, correspondant au cas simple. Et un deuxième jeu de données avec une moyenne des deux groupes assez proches, bien que de variances différentes. Ce deuxième jeu se veut plus représentatif des données Snecma.

Nous avons simulé les points « verts » et « bleus » à l'aide d'une binomiale. La variable qualitative « couleur » peut donc être intégrée au sein de l'équation (5.1) dans l'équation logistique :

$$g_{vk} = \frac{e^{a_v + b_v * Q_k}}{1 + e^{a_v + b_v * Q_k}};$$

avec

$$Q_k = \begin{cases} 1 & \text{si point vert ;} \\ 0 & \text{sinon .} \end{cases}$$

Nous allons donc essayer de retrouver les paramètres du modèle par l'algorithme EM en ignorant la couleur des points, celle-ci n'étant présente que

dans cette équation logistique. Ce faisant, nous sommes supposés retrouver un paramètre  $a_v$  fortement négatif et un paramètre  $b_v$  très élevé (ce qui donne des proportions proche de 1 et 0 en fonction de la valeur de la variable qualitative  $Q_k$ ). Par ailleurs, nous avons initialisé les proportions à 0.5 ( $a_v = b_v = 0$ ), loin de la vraie valeur.

Nous constatons, lorsque nous cherchons les estimateurs des paramètres de notre modèle hiérarchique, que les deux jeux de données mènent à un comportement différent. En effet, nous rappelons que l'algorithme EM utilisé est fortement sensible aux problèmes de maxima locaux multiples. Ainsi, les données A) mènent à une estimation stable des paramètres, des variations de 5% des paramètres par rapport aux vrais paramètres (connus ici) permettent de retrouver les deux groupes.

Le cas des données B) est plus complexe, si le maximum de vraisemblance est bien obtenu avec les bons paramètres, l'estimation est instable. Une faible variation (de l'ordre de 5%) peut entraîner la convergence de l'algorithme vers un maximum local autre que le maximum global. Ce qui se produit généralement est que l'une des classes est totalement vidée, on retombe donc sur notre modèle de mélange « amorçage-propagation » sans groupe. Toutefois, ce n'est pas la seule possibilité et de nombreuses configurations peuvent intervenir avec de mauvaises classifications proposées. Dans le cas où la valeur « bleu-vert » est inconnue, il semble donc difficile de la retrouver.

### Données réelles

Afin de limiter les problèmes de maxima multiples, nous avons réalisé de multiples initialisations et gardé uniquement celle donnant la plus grande vraisemblance. Toutefois, au vu des difficultés observées sur les données simulées, l'atteinte du maximum global n'est pas garantie.

On constate sur le graphique 5.6 que l'ajustement est meilleur dans le cadre de l'utilisation du modèle hiérarchique, c'est particulièrement visible pour la déformation de 0.45. Ainsi à première vue, cette modélisation semble plus adaptée à notre problème. Toutefois, l'algorithme ne retient pas d'information supplémentaires dans la probabilité d'être « vert » ou « bleu », c'est-à-dire que cette probabilité estimée est quasiment constante. Ainsi, bien que la vraisemblance soit plus élevée qu'un modèle 'amorçage-propagation', il n'est pas possible d'en interpréter les résultats en termes d'influence de paramètres de production.

Dans notre modélisation, pour l'amorçage, la moyenne est résumée par une seule fonction linéaire de l'allongement et une variance constante par groupe de données. Or, il est probable que la réalité soit plus complexe que

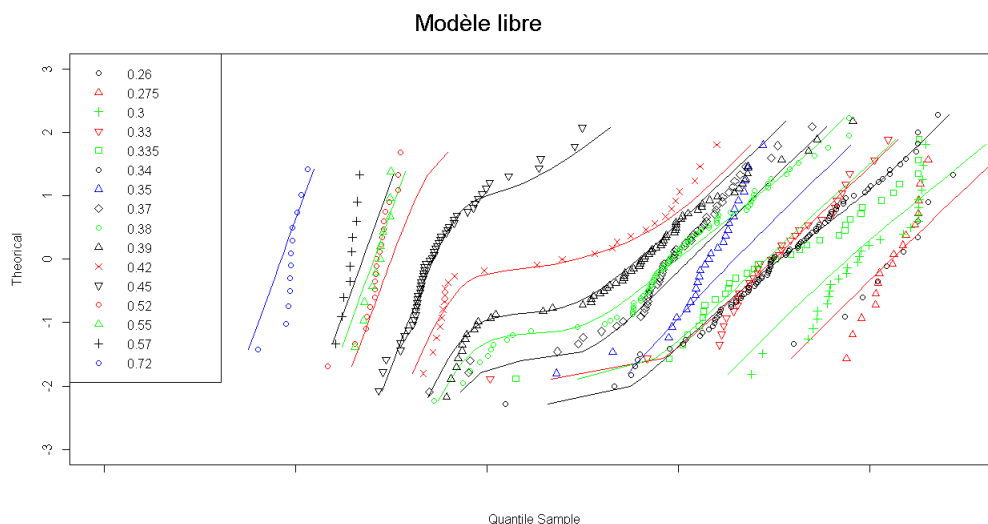


FIGURE 5.6 – Application du modèle hiérarchique à nos données

cette hypothèse simple. Ainsi, notre algorithme s'attache trop à faire coïncider le mélange aux distributions des données. Il est en effet toujours possible d'approcher une distribution quelconque par un modèle de mélange. Et c'est donc probablement ce phénomène qui se produit lorsque l'on utilise les modèles hiérarchiques. Il nous faut dès lors trouver un modèle qui force les groupes à être issus d'une classification des données de production, à la fois pour des questions de stabilité mais aussi d'interprétation des résultats.

## 5.5 Modèle ascendant

Le modèle proposé ci-après mettra en avant une supervision des données, contrairement au modèle non-supervisé précédent. Nous avons donc un jeu de différents paramètres qualitatifs et nous cherchons ceux qui influent sur la durée de vie. Toutefois un modèle additif classique, où la durée de vie serait influencée par une variable moteur et une variable pièce, semble inapproprié. En effet, il y a peu de points communs entre deux pièces d'un même moteur. Et deux pièces ayant la même mission peuvent avoir des caractéristiques complètement différentes. Ainsi notre approche sera de considérer des groupes « pièce + moteur + taille de grain ».

En procédant de la sorte, avec les paramètres d'information sur le moteur, la pièce et la taille de grain, et en supprimant les groupes vides, nous obtenons une subdivision en 10 groupes pour la température de 450°C. Le groupe

ayant le nombre de données le plus faible est composé de 14 points. Nous pouvons donc appliquer notre modèle de mélange amorçage-propagation (4.10) à chacun de ces groupes. Il nous faut définir une stratégie afin de déterminer les groupes semblables et dont les paramètres peuvent être regroupés, des groupes qui doivent rester distincts.

*Remarque :* Le découpage des données a le désavantage de faire décroître rapidement le nombre de données par groupe avec l'augmentation du nombre de variables. C'est pour cela que nous nous limitons dans un premier temps aux trois variables qui donnent le plus d'information sur le matériau, c'est-à-dire le moteur, la pièce et la taille de grain.

Dans un premier temps nous allons chercher à regrouper ensemble les blocs de données les plus semblables. À partir de nos 10 blocs, nous allons former 9 groupes en unissant les 2 blocs dont le regroupement fait le moins varier la vraisemblance globale du modèle (voir Figure 5.7). Nous réitérerons ensuite ce regroupement jusqu'à l'obtention un seul groupe de données. Nous construisons donc un algorithme de regroupement pas à pas visualisable sur la Figure 5.7. Toutefois, le cheminement nous mène à 9 modèles distincts (celui à 10 blocs, à 9  $\dots$ ). Il nous faut donc envisager une stratégie de sélection de modèle afin de déterminer quel(s) modèle(s) utiliser dans cette collection.

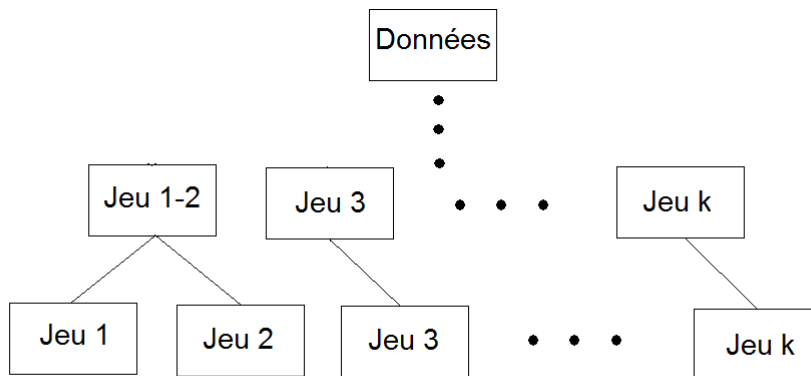


FIGURE 5.7 – Schema du principe des regroupements des données

### 5.5.1 Modèle

Nous noterons  $W = (w_1, \dots, w_N)$  le vecteur indicateur donnant l'appartenance des  $N$  points aux  $K$  groupes  $\{G_1, \dots, G_K\} = \Omega : w_i \in \{1, \dots, K\}$ .

avec  $w_i = k$  si le point  $i$  appartient au  $k$ -ème groupe. La log-vraisemblance du modèle s'écrit :

$$\begin{aligned} \log V(N_i | S_i, w_i, \theta, \alpha, \beta) &= \prod_{k=1}^n \pi(S_k, w_k) f_p(N_k; S_k, \theta_p) \\ &\quad + (1 - \pi(S_k, w_k)) f_p * f_a(N_k; S_k, w_k, \theta), \end{aligned} \quad (5.2)$$

avec  $\theta = (a_{i_1}, b_{i_1}, \sigma_{i_1}, \dots, a_{i_K}, b_{i_K}, \sigma_{i_K})$ . L'expression de l'amorçage devient :

$$\begin{aligned} f_a(N_k; S_k, w_k, \theta) &\sim \mathcal{LN}\left(\sum_j (a_{i_j} + b_{i_j} S_k) 1_{w_k=j}, \sum_j \sigma_{i_j} 1_{w_k=j}\right). \\ \pi(S_k, w_k) &= \frac{e^{\sum_{j=1}^B (\alpha_j + \beta_j S_k) 1_{w_i=j}}}{1 + e^{\sum_{j=1}^B (\alpha_j + \beta_j S_k) 1_{w_i=j}}}. \end{aligned}$$

Les paramètres  $\alpha_j, \beta_j$  ne sont influencés que par les points du groupe  $G_j$ .

### 5.5.2 Algorithme

1. Initialisation :
  - $\Omega^{(0)} = \Omega \quad X^{(0)} = X$
  - Calcul des paramètres du modèle à  $B$  groupes,  $B$  étant le nombre initial de groupes (10 à 450°C) :
 
$$(\hat{\theta}, \hat{\alpha}, \hat{\beta}) = \underset{\theta, \alpha, \beta}{\operatorname{argmax}} \log V(N_i | S_i, w_i, \theta, \alpha, \beta)$$
  - Calcul de la logvraisemblance du modèle :  $\log V^{(0)} = \log V(N_i | S_i, w_i, \hat{\theta}, \hat{\alpha}, \hat{\beta})$
2. Etape Réursive :
  - A l'étape  $k$ , il ne reste que  $B - k$  groupes :

- $$\Omega^{(k)} = \{G_1^{(k)}, \dots, G_{B-k}^{(k)}\}$$
- $\forall c, d \in (1, \dots, B - k), c \neq d$  :
    - On regroupe le bloc  $c$  et le bloc  $d$  ensemble :
    - on obtient l'ensemble  $\Omega_{cd}^{(k)}$  de taille  $B - k - 1$ .
    - Calcul de  $\log V_{cd}^{(k)}$ , la logvraisemblance associée à la nouvelle partition  $\Omega_{cd}^{(k)}$ .
  - On garde le modèle possédant le maximum de logvraisemblance sur tous les modèles testés :

$$\log V^{(k+1)} = \max_{c,d} (\log V_{cd}^{(k)})$$



3. Critère d'arrêt :  $k = B - 1$ . Les données sont toutes regroupées en un seul groupe.

Nous obtenons ainsi un arbre où la racine est notre modèle de mélange « amorçage-propagation ». Il ne reste alors plus qu'à choisir un critère afin de sélectionner le nombre optimal de groupes, ce qui revient en pratique à choisir à quels niveaux couper l'arbre. En effet, l'utilisation de la simple logvraisemblance conduirait à prendre le modèle le plus complexe, c'est-à-dire le modèle avec le maximum de blocs. Une méthode classique de sélection de modèle consiste à créer un critère pénalisé, afin de choisir un modèle plus parcimonieux.

### Pénalisation de la vraisemblance

L'idée d'une pénalisation de la vraisemblance est donc de sélectionner un modèle parmi tous les modèles disponibles.

Lorsque l'échantillon dont on dispose est de très grande taille, une façon simple d'évaluer la qualité d'un modèle parmi plusieurs modèles candidats, est d'utiliser la validation croisée : il s'agit de séparer l'échantillon global en une partie apprentissage (2/3 de l'échantillon global par exemple) et une partie test (les 1/3 restant) servant à l'évaluation (calcul de la somme des carrés des erreurs par exemple). Malheureusement les échantillons sont souvent de tailles réduites, et ce procédé n'est pas toujours applicable. Dans notre cas, il apparaît peu judicieux de réduire le nombre de données globales au vu de la difficulté d'estimation des paramètres.

Un critère pénalisé ajoute à la vraisemblance, une pénalisation de la complexité du modèle. Un critère pénalisé sera de la forme :

$$crit(K) = \log L(\hat{\theta}_K^{MLE}) - pen(K)$$

où  $K$  représente le nombre de paramètres du modèle.

*Remarque :  $pen(K) > 0$  peut dépendre du nombre de données disponibles.*

La première partie du critère représente l'adéquation du modèle. Plus la vraisemblance est grande, meilleure est l'adéquation. La pénalisation a elle pour rôle de contrôler la complexité du modèle.

De nombreux critères de vraisemblance pénalisée existent. On peut citer l'AIC (Akaike's Information Criterion [3]) ou le BIC (Bayesian's Information Criterion)(Schwarz [40]). Ces deux critères sont issus de considérations asymptotiques. Le critère BIC est basé sur la maximisation de la vraisemblance intégrée :

$$f(y|K, m) = \int f(y|K, m, \theta)l(\theta|K, m)d\theta,$$

où  $l(\theta|K, m)$  est la distribution a priori du vecteur des paramètres. Cette vraisemblance intégrée est difficilement calculable, une approximation de Laplace est donc utilisée. Finalement, le modèle sélectionné minimise le critère BIC défini par :

$$BIC(K, m) = -\log[f(y|(K, m, \hat{\theta}))] + \frac{\lambda_{(K,m)}}{2} \log(N),$$

où  $\lambda_{(K,m)}$  dénote le nombre de paramètres libres de la collection  $(K, m)$ .

### 5.5.3 Stabilisation de l'algorithme

Le principal défaut de cet algorithme est qu'il demande l'acquisition de nombreux paramètres (8) pour peu de données lors des premières itérations. Le découpage en blocs crée en effet des jeux de données d'un minimum de 15 points. Nous sommes donc en deçà de ce qui est préconisé pour une bonne estimation (voir section 4.4.3). Afin de pallier cet inconvénient, nous recommandons de fixer les paramètres suivants aux valeurs obtenues avec le modèle appliqué sur toutes les données :

- les estimations des paramètres de propagation  $\theta_p$ . En effet la variation moindre des paramètres de propagation ainsi que la bonne estimation de la courbe en propagation pure nous permet de fixer ces paramètres,
- les paramètres de la logistique sont communs aux groupes et fixés. En effet, les paramètres de la logistique dépendent fortement des niveaux de déformation testés. Or, la base de donnée est fortement déséquilibrée. Ainsi garder des paramètres uniques pour la logistique permet de stabiliser l'algorithme tout en atténuant l'effet du déséquilibre du plan d'expérience sur les résultats. On a donc  $\alpha_1 = \alpha, \dots, \alpha_K = \alpha$  et  $\beta_1 = \beta, \dots, \beta_K = \beta$ .

La logvraisemblance des différents modèles reste calculée à l'aide de la formule (6.1). Nous utiliserons le critère BIC pour la comparaison de tous ces modèles, avec dans notre cas, la pénalisation suivante :  $\lambda(K, m) = K * 3 + 5$ .

- 3, représente le nombre de paramètres d'amorçage pour un bloc. On le multiplie donc par  $K$ , le nombre de blocs.
- $5 = 3 + 2$ , représente les 3 paramètres de propagation fixés pour tous les groupes et les 2 paramètres de la logistique.

Dans la section suivante, nous discuterons l'utilisation de ces approximations, notamment le blocage des paramètres de la logistique.

### 5.5.4 Résultats sur données réelles

#### Température d'essai de 450°C

La Figure 5.8 retrace l'évolution de notre critère pénalisé en fonction de l'évolution du nombre de groupes de notre modèle. Le but étant la minimisation de ce critère, on constate que ce minimum est atteint pour 2 groupes. Ceux-ci sont visibles avec l'arbre de la Figure 5.9. On constate que la première division est liée à la taille de grain. Le premier facteur discriminant est donc la taille de grain. Ainsi la durée de vie en amorçage pour une taille de grain petite sera en moyenne plus longue que pour une taille de grain plus grosse. Selon le critère BIC, le modèle retenu doit donc distinguer les tailles de grain. Toutefois si on regarde quel découpage est ensuite réalisé, on retrouve un découpage au sein des petits grains. Le moteur D et l'anneau attendant du moteur E (moteur à forte puissance) surpassent en moyenne les autres moteurs proposant la même taille de grain. Bien que ce modèle ne soit pas retenu, on constate que l'arbre est cohérent, puisque la fabrication du moteur D repose sur des normes matériau plus exigeantes qu'un moteur civil. On retrouve donc une trace de cette exigence dans nos données.

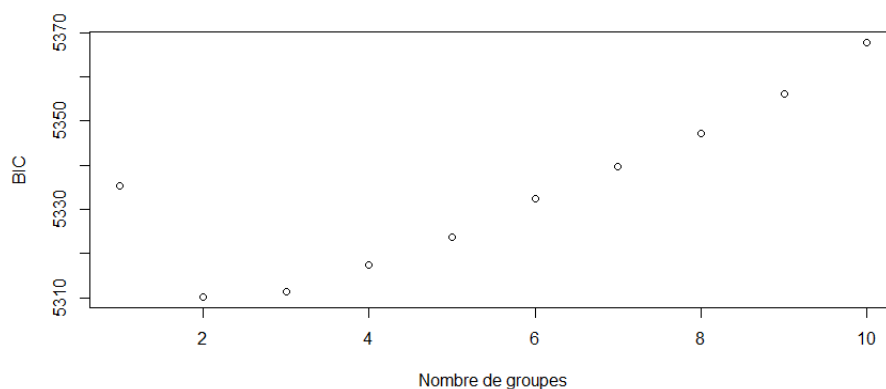


FIGURE 5.8 – Evolution du critère BIC en fonction du nombre de groupes, paramètre de logistique fixés

Toutefois une question peut être posée sur nos résultats. Nous avons en effet utilisé des valeurs contraintes sur les paramètres de propagation et de logistique. N'influent-elles pas sur les regroupements ? Pour répondre à cette question nous allons réutiliser le même algorithme. Cette fois-ci, les paramètres de la logistique seront différents pour chaque groupe. Cela nécessite un ajustement du critère pénalisé qui devient :

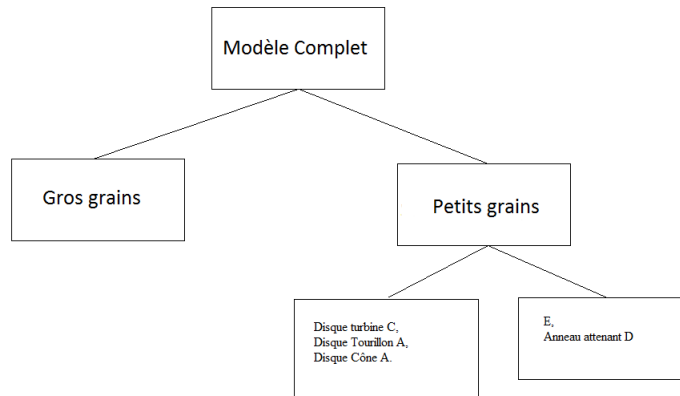


FIGURE 5.9 – Arbre obtenu pour 3 groupes

$$BIC(K, m) = -\log[f(y|(K, m, \hat{\theta}))] + \frac{\lambda_{(K,m)}}{2} \log(N)$$

avec  $\lambda_{(K,m)} = K \times 5 + 3$ , puisque les 2 paramètres précédemment fixés sont libres. Le critère est visualisable sur la figure (5.10)

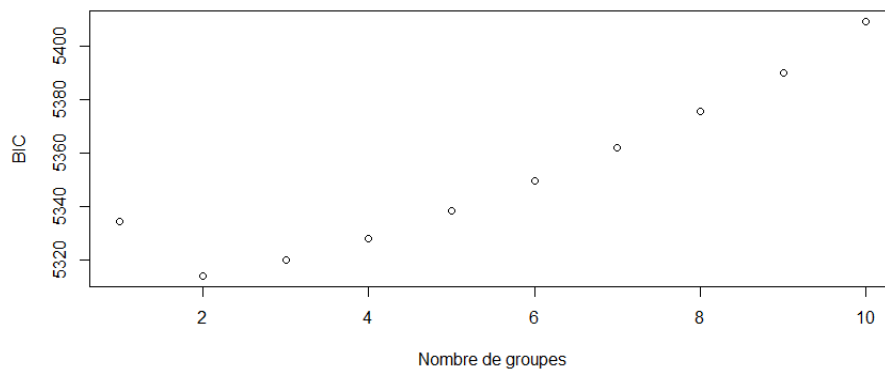


FIGURE 5.10 – Évolution du critère BIC en fonction du nombre de groupes, paramètres de logistique libres

Plusieurs remarques peuvent être effectuées :

- l'allure de la courbe est similaire à celle obtenue en fixant les paramètres de la logistique,
- le minimum du critère BIC est obtenu pour 2 groupes comme précédemment,

- à nombre de groupes équivalent, le modèle avec paramètres de logistique libres possède un critère BIC plus élevé.

Cela signifie que la méthodologie employée, et notamment le fait de fixer les paramètres de la logistique semble cohérent, puisque l'information apportée par ceux-ci n'est pas suffisante et est rejetée par notre critère BIC.

Par ailleurs, le découpage a une signification du point de vue du matériau. En effet de nombreux modèles de mécanique des matériaux incluent la taille de grains dans leur modélisation (voir Tomkins [47] par exemple). On retrouve donc l'importance de ce paramètre dans la durée de vie en amorçage de nos éprouvettes.

### Température d'essai de 550°C

La figure (5.11) retrace l'évolution du critère BIC pour une température de 550°C en fonction du nombre de groupes.

*Remarque : Le nombre maximum de groupes est en fait de 15, mais la courbe croît par la suite et nous nous contentons ici d'un point de vue local.*

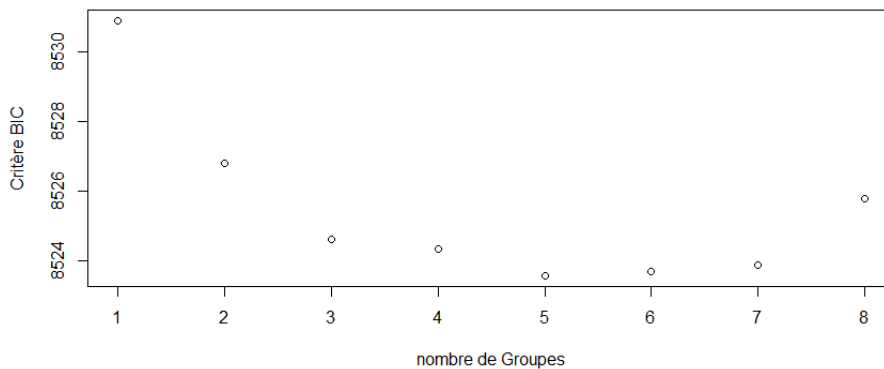


FIGURE 5.11 – Evolution du critère BIC en fonction du nombre de groupes, paramètres de logistique libre

Composition des 5 groupes retenus :

1. Gros grains : B - Disque Cône D - Disque Turbine C.  
Petits grains : Disque Turbine A - Disque Laby A
2. Gros grains : A divers.  
Petits Grains : Disque Turbine C - A Divers
3. Gros grains : Disque Laby C - Dique Laby A - Disque Turbine A

4. Petits Grains : B - Disque Cône D - Disque laby C.
5. Petits Grains : Anneau D.gm

Le nombre de groupes retenus est bien plus important qu'à 450°C,. Si pour les 3 derniers groupes on constate une distinction gros/petits grains avec une durée de vie en moyenne plus élevée des éprouvettes issues du MFP-Anneau (tout comme à 450°C). Le phénomène est plus difficile à décrypter pour les deux premiers groupes (composés de gros et de petits grains). On peut toutefois noter que le premier groupe est composé principalement de gros grains avec l'apparition d'échantillons issus de petits grains du moteur A. Le deuxième groupe est composé de petits grains, mais l'on retrouve là aussi le disque A. La multiplication des groupes peut en partie s'expliquer par l'augmentation, avec la température, de l'influence de l'oxydation dans le phénomène de fatigue, comme noté par Fournier et Pineau [22]. Il semble toutefois difficile de s'avancer plus dans les suppositions sans élément supplémentaire, d'autant que les pièces testées ne sont pas les mêmes aux deux températures.

### Critique du modèle

Si ce modèle donne des résultats probants à 450°C, la multiplication des groupes à 550°C sans que les ingénieurs matériau ne puissent formuler d'hypothèses quant aux rattachements des différents groupes peut laisser dubitatif. Toutefois une explication peut être fournie quand à la difficulté d'obtenir des groupes cohérents et ayant une signification métallurgique précise. En effet un phénomène peut en partie fausser nos résultats. Il s'agit du déséquilibre de la base de données.

Un exemple est donné avec le graphique (5.12). Il retrace l'appartenance des points sur les données classifiées en amorçage. On observe en rouge les points du disque de turbine du A.

Plusieurs remarques peuvent ainsi être faites :

- on voit sur le graphique que les points des deux groupes ne sont pas forcément testés aux mêmes déformations. Notamment dans la partie des faibles déformations,
- le disque de Turbine du A ne possède que peu de niveaux testés. Ce qui rend difficile l'estimation d'une pente pour ces points,
- par ailleurs ce moteur est testé à des niveaux de déformation où les courbes pour les gros grains et les petits grains sont relativement proches, rendant d'autant plus difficile leur classification dans un des deux groupes.

On observe donc par ce biais le déséquilibre de la base de données qui complique l'analyse des résultats.

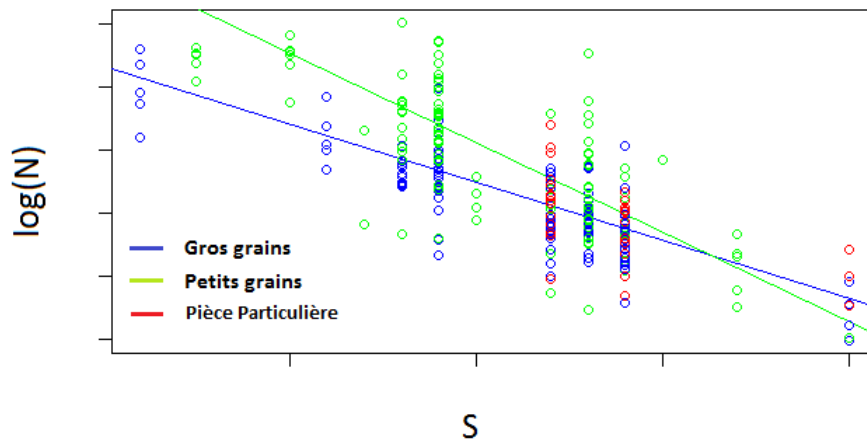


FIGURE 5.12 – Visualisation des groupes sur les données classifiées en amorçage à 450°C.

Ce modèle a toutefois été construit dans une optique précise, expliquer les comportements observés sur le graphique (5.13) où les courbes, dans la partie amorçage, se croisent à des niveaux de déformation différents. Nous pouvons donc maintenant créer deux graphiques, un pour les petits grains et un pour les gros grains. Par ailleurs, nous visualiserons uniquement les données classifiées en mode 2 afin de limiter la superposition d'information sur les graphiques.

Plusieurs remarques peuvent être effectuées :

- le découpage par groupe semble être cohérent, on observe bien une durée de vie et une variance plus élevée dans le groupe 2 que dans le groupe 1,
- bien que des croisements subsistent encore, par exemple dans le groupe 1 entre les tests effectués à une déformation de 0.33 et ceux à 0.34, les groupes sont plus homogènes que lorsqu'une unique distribution en amorçage est présente. Il reste toutefois des informations que nous n'avons pas exploité, le nombre de groupes étant limité pour des raisons de lenteur de l'algorithme. On peut citer par exemple la zone de prélèvement au sein de la pièce et dont les caractéristiques peuvent varier en fonction de la distance à la surface de la pièce.

Ainsi, le modèle remplit son rôle qui était de donner une explication aux chevauchements des courbes sur le « quantile plot », le principal facteur étant

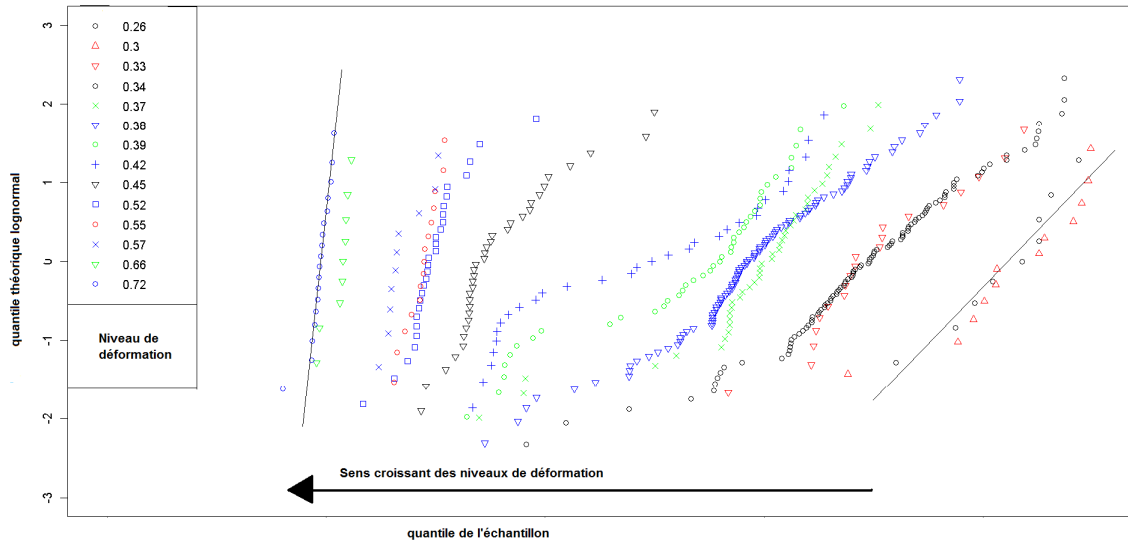


FIGURE 5.13 – « quantile plot » des données à 450°C

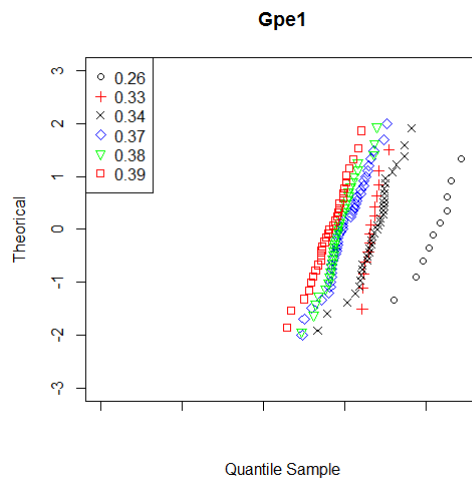


FIGURE 5.14 – « quantile plot » des données en amorçage à 450°C - groupe 1

étant l'influence de la taille de grain sur la durée de vie en fatigue.



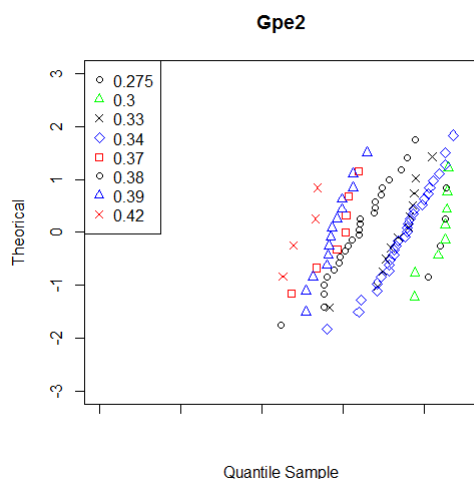


FIGURE 5.15 – « quantile plot » des données en amorçage à 450°C - Groupe 2

## 5.6 Conclusion

Nous avons donc testé deux méthodes de prise en compte d'informations de production. Informations qui ne sont qu'un moyen détourné d'atteindre les vrais paramètres physiques qui pourraient expliquer la différence de comportement, telles que les différentes températures et pressions obtenus localement pendant le forgeage. Le modèle hiérarchique proposant une méthode non-supervisée, échoue à extraire l'information pertinente de ces données de production. Une partie de l'explication de ce mauvais comportement peut certainement être donnée par l'écart entre l'expérience et la modélisation assez simple que nous en avons réalisée. C'est ce biais qui est compensé par l'utilisation du modèle hiérarchique mais, dès lors, ce modèle s'éloigne de l'objectif initial, à savoir trouver les paramètres influents dans la durée de vie. Son utilisation ne semble donc pas pertinente pour notre problématique.

Le modèle descendant a donc été construit pour remédier à ce phénomène. Il est construit de telle sorte que retenir plusieurs groupes revienne à différencier les données en fonction des informations sur le matériau. Il est toutefois confronté à des limites intrinsèques :

- l'algorithme de recherche des groupes ne permet pas de multiplier le nombre de variables étudiées. Le modèle est donc inutilisable pour rechercher de manière exploratoire les variables influentes de production. Toutefois, sa facilité d'interprétation peut permettre de rapidement vérifier des hypothèses sur un nombre faible de variables. Il peut donc

être utilisé dans un cadre de validation d'hypothèses fournies par des ingénieurs matériau.

- le deuxième écueil tient au fait que les groupes obtenus ne sont pas forcément interprétables. En effet si l'étude à 450°C donne des résultats facilement compréhensibles, l'étude des données à 550°C est difficilement exploitable telle quelle. Cette méthodologie ne peut donc pas être généralisée et nécessite donc le point de vue d'un ingénieur matériau quand aux conclusions que l'on peut en tirer.
-

# Chapitre 6

## Conclusion et perspectives

### 6.1 Conclusion

Au travers du travail présenté dans ce mémoire, nous nous sommes d'abord intéressé à offrir une modélisation statistique qui ait du sens du point de vue de la rupture mécanique des matériaux. Tout en gardant à l'esprit de créer un modèle qui représente fidèlement la dispersion observée sur les données. Ceci nous a mené à un modèle de mélange non-classique :

$$\pi(S)f_{N_i} + (1 - \pi(S))f_{N_i+N_p}$$

Plusieurs difficultés se sont donc présentées, telles que l'utilisation de somme de lognormales mais aussi l'utilisation de la variable  $S$  à la fois en tant que régresseur pour la moyenne des lognormales, mais aussi en tant que variable active pour le calcul des proportions  $\pi(S)$  de ce mélange. Les propriétés théoriques de ce modèle tels que l'identifiabilité ont été validées et les résultats obtenus sont satisfaisants que ce soit du point de vue des simulations où de l'application sur données réelles.

Ce modèle a par ailleurs été étendu afin de pouvoir intégrer des informations supplémentaires qui peuvent être issues, soit de la production, soit des essais.

### 6.2 Perspectives

Les perspectives d'amélioration du modèle sont nombreuses. Elles reposent pour la plupart sur une augmentation de l'information disponible :

- L'augmentation du nombre de points de la base de données et l'équilibrage de celle-ci. Afin notamment d'améliorer la précision dans le calcul des proportions de points en propagation ou en amorçage + propagation.
-

- L'augmentation du nombre de variables étudiées, voire la modification de certains tests de fatigue dans le but de déterminer l'effet de la taille de l'éprouvette sur la durée de vie. Il convient donc de veiller à équilibrer la base de données au cours des prochaines années.

### 6.2.1 Equilibrage du plan d'expérience

Le principal écueil rencontré au cours de cette thèse, notamment avec l'intégration des données de production est le déséquilibre de la base de données. En effet une pièce de moteur est rarement testée à plus de 6 niveaux de déformation, là où une température compte une vingtaine de niveaux testés.

#### Quels niveaux choisis ?

Lorsque l'on observe la courbe de Wöhler issue des données Snecma, on se rend compte que l'on peut distinguer 3 zones avec des nombres d'essais différents (voir graphique 6.1).

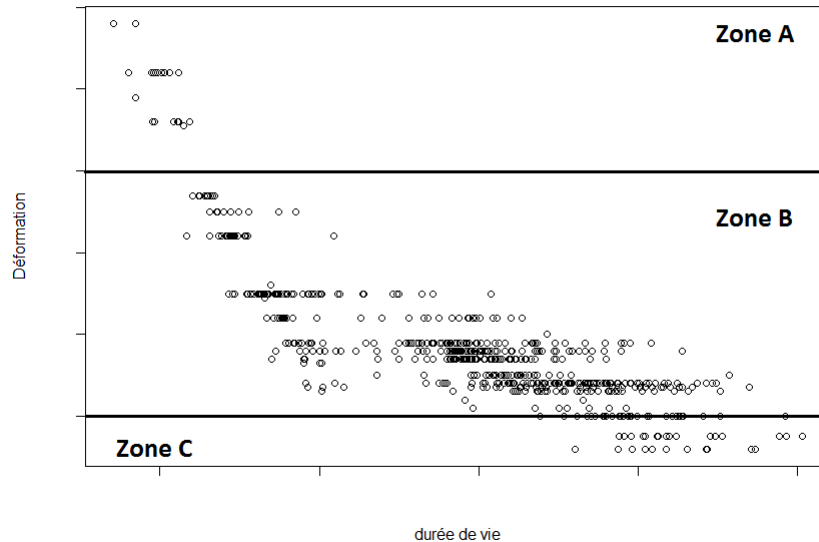


FIGURE 6.1 – 450°C, visualisation de la courbe S-N issue des données snecma.

- La Zone A ne présente que peu d'intérêt. Elle est en effet située au-dessus de la zone d'utilisation du moteur et n'offre que peu de disper-

---

sions. Il ne semble donc pas nécessaire de rajouter quantité de points dans cette zone.

- La Zone B est la zone principale où le plus d'éprouvettes sont situées. Elle est primordiale pour plusieurs raisons. C'est d'abord la zone utile d'utilisation en vol, c'est donc à partir de ces niveaux de déformations qu'est défini le dimensionnement de la pièce. Par ailleurs, il s'agit de la zone où apparaît le phénomène de bi-modalité. La précision des estimations de notre modèle dépend fortement du nombre de points présents dans cette zone.
- La zone C est la zone des très fortes durées de vie. L'augmentation du nombre de points dans cette zone peut se faire si l'on s'intéresse aux comportements du matériau pour les très faibles contraintes. Il ne s'agit pas a priori d'une zone sensible pour le dimensionnement. (voir section 6.1.3, pour la limite d'endurance).

Nous voyons donc que l'attention des futurs essais doit se porter particulièrement sur la Zone B du graphique, ce qui était déjà naturellement réalisé par les ingénieurs de Snecma-Safran. Toutefois, il convient, au sein de cette zone de veiller tout particulièrement à ce que les moteurs soient représentés sur plusieurs niveaux de contraintes.

### Intégration de modèles de mécanique des matériaux

L'ingénieur en mécanique des matériaux peut être désorienté par l'approche fortement statistique de cette thèse, et notamment, par la non-exploitation de modèles mécaniques. Nous allons tenter d'en donner l'explication ici, et ce que nécessite de notre point de vue l'intégration de plusieurs paramètres matériaux supplémentaires.

**Modèle multi-température** Un modèle proposant l'intégration d'un paramètre de température n'offrirait pas de garanties de prédiction en l'état actuel de la base de données. Nous avons en effet vu au chapitre 4 que les 4 températures testées (200, 350, 450 et 550°C) peuvent en effet être regroupées, a minima, en deux classes. Les données à 200 et 350°C et celles à 450 et 550°C. Une modélisation naïve à base de relations linéaires reposerait donc sur 2 modèles ne possédant chacun que deux températures de référence. La prédiction à d'autres températures que celles étudiées s'avèrerait donc de piètre qualité avec aucune garantie que le comportement au sein de chaque groupe soit linéaire, et donc, aucune garantie que la prédiction à d'autres températures soit bonne. Seuls l'ajout de points à des températures intermédiaires peut amener une meilleure prise en compte de la température dans notre modèle. Cependant, au vu de la complexité actuelle du modèle,

---

l'ajout d'un nouveau paramètre nécessitera un nombre de points supplémentaires conséquent aux diverses températures.

### Comportement du modèle pour les très longues durées de vies

Le sujet des très longues durées de vie est délicat. Le coût d'un essai de fatigue fait que l'éprouvette peut ne pas être rompue à la fin de l'essai. Il est classique de considérer une limite d'endurance au-dessous de laquelle le matériau ne rompt pas. Cette limite pouvant être fixe ou aléatoire. Notre modèle n'intègre pas une telle limite. La première raison est qu'elle n'apparaît pas clairement dans notre jeu de données. En effet en utilisant le modèle statistique le plus adapté à une limite d'endurance aléatoire (celui de Pascual-Meeker [34]) :

$\log N = a + b \log(S - S'_0)$  ;  $S'_0 \sim \mathcal{N}(\alpha, \beta)$  et désignant la limite d'endurance.

Et en l'appliquant à nos données en « amorçage + propagation » (puisque ce sont ces données qui sont concernées par cette limite). On obtient le graphique 6.2.

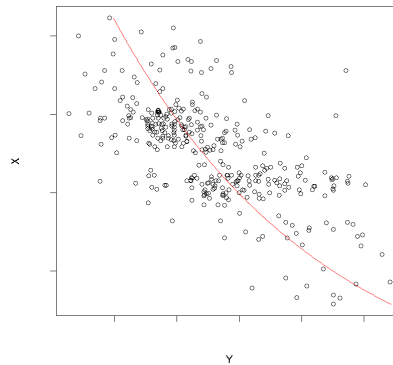


FIGURE 6.2 – Modèle de Pascual-Meeker appliqué aux données classées en « propagation + amorçage »

Si ce dernier diffère légèrement de la droite que nous proposons comme modélisation, l'estimation de la limite d'endurance est proche de 0, la forme de la courbe n'étant donnée que par la forme logarithmique de la fonction.

Ainsi il n'est pas évident que le rajout d'une limite d'endurance soit pertinent dans le cadre de nos données. On pourra citer à ce sujet l'article 'Is there a fatigue limit' (Pittel et al. [35]) qui montre que le sujet est complexe. Toutefois, son rajout dans notre modèle reste possible si l'incorporation de données aux faibles déformation indique un tel comportement.

Il suffira de remplacer la distribution lognormale  $f_{N_i}$  dans le modèle suivant :

$$f_N = \pi(S)f_{N_p} + (1 - \pi(S))f_{N_i+N_p} \quad (6.1)$$

par la distribution du modèle de Pascual-Meeker.

La stabilité de l'algorithme résultant devra être étudié avec précaution puisque tant l'algorithme de Pascual-Meeker que celui du modèle de mélange « amorçage-propagation » nécessitent d'importants moyens de calcul numériques. Notamment les deux modèles nécessitent l'approximation d'intégrales.

### Effet d'échelle

Nous nous sommes intéressé, au cours de cette thèse, uniquement à la durée de vie en fatigue des éprouvettes. Or celles-ci ne sont finalement qu'un moyen d'aboutir à une estimation de la durée de vie complète d'une pièce du moteur. Notre modèle étant principalement statistique, il ne rend pas compte de ce que peut être la fatigue pour des échantillons de matériau de volumes différents des éprouvettes. Ainsi, les quantiles donnés sont-ils ceux applicables uniquement à des volumes de taille comparable. Si les zones de matériau fortement sollicitées d'une pièce ne représentent pas tout le volume de la pièce, il conviendrait d'avoir des jeux d'éprouvettes plus volumineuses afin de pouvoir quantifier l'effet d'une variation d'échelle. Cela permettrait d'estimer l'influence d'une variation du volume et du rapport  $\frac{Volume}{Surface}$  et donc se rapprocher des gros volumes. En effet si on reprend les travaux de Bussac et Lautridou [8], on s'aperçoit que la nocivité d'un défaut est supérieure en surface et subsurface qu'en interne. L'augmentation du volume pose quand à elle la question de la probabilité de présence d'un défaut suffisamment nocif pour amorcer rapidement.

La première étape de cette modélisation serait d'intégrer le volume dans le calcul de la probabilité d'appartenir au groupe d'amorçage rapide, en remplaçant  $\pi(S)$  par  $\pi(S, V) = \frac{e^{\alpha+\beta S+\gamma V}}{1+\alpha+\beta S+\gamma V}$  dans le modèle 4.10 :

$$f_N = \pi(S, V)f_{N_p} + (1 - \pi(S, V))f_{N_i+N_p}$$

Cette méthode a l'avantage de la simplicité puisque le rajout d'information dans la partie logistique du modèle n'alourdit pas (ou très peu) le calcul des estimateurs, grâce à l'algorithme EM qui dissocie leurs calculs de celui des autres paramètres.

Un autre facteur à prendre en compte pour le passage de l'éprouvette à la pièce est le type de chargement. Ici nous sommes sur un chargement

uniaxial cyclique imposé, alors que la réalité est certainement plus proche d'un chargement multiaxial (plus ou moins aléatoire). Toutefois, la prise en compte de ce type de chargement est bien plus complexe dans le cadre de notre modèle que l'introduction d'une variation de volume de l'éprouvette.

---



---

## Bibliographie

- [1] Abramowitz, M., Stegun, I. A. (1972), Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, *New York : Dover Publications*
  - [2] Agnoli, A., Bernacki, M., Logé, R., Franchet, J.M., Laigo, J., Bozzolo, N., (2012). Understanding and Modeling of Grain Boundary Pinning in Inconel 718. *Superalloys 2012*
  - [3] Akaike, H., (1973), A new look at the statistical model identification, *IEEE Transactions on Automatic Control* 6, 716-723.
  - [4] Alexandre, F. (2004) Aspects probabilistes et microstructuraux de l'amorçage des fissures de fatigue dans l'alliage INCO 718. *Phd thesis*. Ecole Nationale Supérieure des Mines de Paris.
  - [5] Asmussen, S. and Rojas-Nandayapa, L. (2005), Sums of dependent lognormal random variables : asymptotics and simulation. *Research report*. Thiele Centre for Applied Mathematics in Natural Science
  - [6] Bathias, C., Pineau, A. (2008-2009) Fatigue des Matériaux et des Structures 1-3 *Hermès science publication*
  - [7] Biernacki, C., Celeux, G., Govaert, G. (2003). Choosing starting values for the EM algorithm for getting the highest likelihood in multivariate Gaussian mixture models. *Computational Statistics and Data Analysis* 41,567-575
  - [8] Bussac, A., Lautridou, J.C. (1993) A probabilistic model for prediction of LCF surface crack initiation in PM alloys. *Fatigue and Fracture of Engineering Materials and Structures*, 8, 861-874
  - [9] Bomas, H., Burkart, K., Zoch, H.W., (2011), Evaluations of S-N curves with more than one failure mode *International Journal of Fatigue*, 33, 19-32.
  - [10] Brand, A., Flavenot, J.P., Grégoire, R., Tournier, C., (1999) Données technologiques sur la fatigue. *Senlis SETIM*
  - [11] Breiman, L, Friedman, J., Stone, C., Olshen, R.A., Classification and regression trees. *s. Monterey, Calif., U.S.A. : Wadsworth, Inc*
-

- [12] Castillo, E., Canteli, A.F., (2009) A unified statistical methodology for modeling fatigue damage. *Springer*
  - [13] Celeux, G. and Diebolt, J. (1985) The sem algorithm : a probabilistic teacher algorithm derived from the em algorithm for the mixture problem. *Computational Statistics Quaterly*, 2, 73-85.
  - [14] Chan, K.W. (2010), Role of microstructure in fatigue crack initiation. *International Journal of Fatigue*, 32, 1428-1447.
  - [15] Chistyakov, V.P., (1964) A theorem of sums of independent positive random variables and its applications to branching processes. *Theory of Probability and Its Applications*, 9, 640-648
  - [16] Crowder, M. (1999) Classical Competing Risks. *Harlow, England : Addison-Wesley*
  - [17] Dempster, A.P., Laird N.M. and Rubin, D.B. (1977) Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society*, 39, 1-38.
  - [18] DeSarbo, W.S. and Cron, W.L. (1988) A maximum likelihood methodology for clusterwise linear regression. *International Journal of Classification*, 5, 249-282.
  - [19] Embrechts, P., Goldie, C.M., (1980) On closure and factorization properties of subexponential distributions. *Journal of the Australian Mathematical Society*, 29, 243-256
  - [20] Fenton, L. F. (1960) The sum of log-normal probability distributions in scatter transmission systems. *IRE Transactions on Communication Systems*, 8, 57-67
  - [21] Fouchereau, R. (2009) Analyse de la tenue en fatigue des pièces en inconel 718. *Rapport de stage Snecma*
  - [22] Fournier, D., Pineau,A. (1977) Low cycle fatigue behavior of inconel 718 at 298 K and 823 K. *Metallurgical Transactions A*. 8, 1095-1105
  - [23] Harlow, D.G., Wei, R., Sakai,T. and Oguma, N. (2006), Crack growth based probability modeling of S-N response for high strength stell. *International Journal of Fatigue*, 28 , 1479-1485.
  - [24] Hanaki, S., Yamashita, M., Uchida, H. and Zako, (2010) M. On stochastic evaluation of SN data based on fatigue strength distribution. *International Journal of Fatigue*, 32, 605-609.
  - [25] Jha, S.K. and Ravi Chandran, K.S. (2003), An unusual fatigue phenomenon : duality of the SN fatigue curve in the b-titanium alloy Ti10V2Fe3Al. *Scripta Materialia*, 48, 1207-1212.
-

- 
- [26] Jha, S.K., Caton, M.J. and Larsen, J.M. (2007), A new paradigm of fatigue variability behavior and implications for life prediction. *Material Science and Engineering :A*, 468-470, 23-32.
- [27] Jha, S.K., Larsen, J.M. and A.H. Rosenberg (2009), Toward a physics-based description of fatigue behaviour in probabilistic life-prediction. *Engineering Fracture Mechanics*, 76, 681-694.
- [28] Jordan, M.I., Jacobs, R.A. (1993) Hierarchical mixture of experts and the EM algorithm. *Neural Computation* 6, 181-214
- [29] Le Biavant-Guerrier, K.(2000) Etude de l'amorçage de fissures de fatigue dans le Ti-6Al-4V, *phd thesis* Ecole centrale Paris.
- [30] Li, X. (2008) A Novel Accurate Approximation Method of Lognormal Sum Random Variables. *Master Thesis* Wright State University.
- [31] McLachlan, G. and Peel, D. (2000), Finite mixture models, Wiley.
- [32] Nelson , W. (1982), Applied life data analysis, Wiley.
- [33] Paris, P. and Erdogan, F., (1963), A critical analysis of crack propagation laws, *Journal of Basic Engineering, Transactions of the American Society of Mechanical Engineers*, 85, 528-534.
- [34] Pascual, F.G. and Meeker, W.Q. (1999), Estimating Fatigue Curves with the Random Fatigue-Limit Model. *Technometrics*, 41, 277-290.
- [35] Pittel, B., Schwerdt, D., Bergerr, C. (2011), Very high cycle fatigue : Is there a fatigue limit ? *International Journal of fatigue*, 33, 49-58.
- [36] Pollak, R.D, Palazotto, A.N., (2009) A comparison of maximum likelihood models for fatigue strength characterization in materials exhibiting a fatigue limit *Probabilistic Engineering Mechanics*, 24, 236-241
- [37] Ravi Chandran, K.S., Chang, P. and Cashman, G.T. (2010) Competing failure modes and complex SN curves in fatigue of structural materials *International Journal of Fatigue*, 32, 482-491.
- [38] Sakai, T., Nakayasu, H. and Nishikawa, I. (2005) Establishment of JSMS standard regression method of SN curves for metallic materials. *Safety and Reliability of Engineering Systems and Structures*, 643
- [39] Sakai, Y., Lian,B., Takeda, M., Shiozawa, K., Oguma, N., Ochi, Y., Nakajima,M., Nakamura, T. (2010) Statistical duplex S-N characteristics of high carbon chromium bearing in rotating bending in very high cycle regime *International Journal of Fatigue*, 32, 497-504.
- [40] Schwarz, G., (1978) Estimating the dimension of a model. *Annals of statistics* 6, 461-464
-

- [41] Schwartz, S., Yeh, Y.S. (1982), On the distribution function and moments of power sums with log-normal components. *Bell System Tech Journal*, 61, 1441-1462
  - [42] Sims, C.Y., Stoloff, N.F., and Hagell W.C., (1987) Superalloy II : High temperature materials for aerospace and industrial power, Wiley.
  - [43] Shiozawa, K., Murai, M., Shimatani, Y. and Yoshimoto, T. (2010) Transition of fatigue failure mode of NiCrMo low-alloy steel in very high cycle regime. *International Journal of Fatigue*, 32, 541-550.
  - [44] Suresh, S., (1998) Fatigue of Materials, *Cambridge University Press*.
  - [45] Szyszkowicz, S.S., Yanikomeroğlu, H. (2009) Fitting the modified-power-lognormal to the sum of independent lognormals distribution. *Proceedings of the 28th IEEE conference on Global telecommunications* 5823-5828
  - [46] Thieulot-Laure, E., (2008) Méthode probabiliste unifiée pour la prédiction du risque de rupture en fatigue *Phd thesis*, ENS Cachan.
  - [47] Tomkins, B., (1968) Fatigue crack propagation - An analysis *Philosophical magazine*, 18, 1041-1066
  - [48] Weixing, Y., Sheijieng, G., (2007) VHCF test and life distribution of aluminium alloy LC4CS. *International Journal of Fatigue*, 32, 497-504.
  - [49] Wu, J., Mehta, N., Zhang, J. (2005) A Flexible Lognormal Sum Approximation Method. *IEEE Global Telecommunications Conference*, 6, 3413-3417
-

---

## Annexe A

# Identifiabilité du modèle de mélange Amorçage Propagation

Cette annexe a pour but la démonstration de l'identifiabilité du modèle suivant.

$$f_N = \pi(s) f_{N_p} + (1 - \pi(s)) f_{N_i + N_p},$$

Sous l'assertion réaliste que  $\sigma_i > \sigma_p$ . Il est difficile de démontrer directement cette identifiabilité, la démonstration sera donc conduite pas à pas.

1. Tout d'abord, nous poserons que  $\pi$  et le niveau de contrainte  $s$  sont fixés. L'unicité des paramètres de dispersion et de position  $\mu$  et  $\sigma$  sera démontrée pour les deux distributions log-normales.
2. Dans un deuxième temps nous autoriserons la variation du niveau de contrainte  $s$ , nous prouverons l'unicité des paramètres  $a$  et  $b$  définissant  $\mu$  ( $\mu = a + bs$ ). De la même manière, nous montrerons l'unicité des paramètres  $\alpha$  et  $\beta$ , paramètres entrant dans le modèle de régression logistique de la proportion  $\pi$ .

**1 :** Soient  $s$  et  $\pi$  fixés. Sous l'hypothèse  $\sigma_i > \sigma_p$ , nous avons l'équivalence asymptotique suivante pour les queues de distribution du modèle (4.4), cf. [5] :

$$\mathbb{P}(N > x) \sim \frac{\pi_2 \sigma_i e^{-\frac{1}{2} \frac{(\log(x) - \mu_i)^2}{\sigma_i^2}}}{\sqrt{2\pi} (\log(x) - \mu_i)}.$$

Donc, en prenant deux paramétrisations pour  $N$ , nous avons :  $(\pi_1, \pi_2 = 1 - \pi_1, \mu_i, \sigma_i, \mu_p, \sigma_p)$  and  $(\pi'_1, \pi'_2 = 1 - \pi'_1, \mu'_i, \sigma'_i, \mu'_p, \sigma'_p)$ ,

---

$$\lim_{x \rightarrow \infty} \left( \frac{\frac{\pi_2 \sigma_i}{\sqrt{2\pi}(\log(x) - \mu_i)} e^{-\frac{1}{2} \frac{(\log(x) - \mu_i)^2}{\sigma_i^2}}}{\frac{\pi'_2 \sigma'_i}{\sqrt{2\pi}(\log(x) - \mu'_i)} e^{-\frac{1}{2} \frac{(\log(x) - \mu'_i)^2}{\sigma_i'^2}}} \right) = 1.$$

C'est à dire

$$\lim_{x \rightarrow \infty} \left( \frac{\pi_2 \sigma_i \log(x) - \mu_i}{\pi'_2 \sigma'_i \log(x) - \mu'_i} e^{-\frac{1}{2} \log^2(x) \left( \frac{1}{\sigma_i^2} - \frac{1}{\sigma_i'^2} \right) - \frac{1}{2} \log(x) \left( -\frac{2\mu_i}{\sigma_i^2} + \frac{2\mu'_i}{\sigma_i'^2} \right) - \frac{1}{2} \left( \frac{\mu_i^2}{\sigma_i^2} - \frac{\mu_i'^2}{\sigma_i'^2} \right)} \right) = 1$$

Et comme  $\lim_{x \rightarrow \infty} \left( \frac{\log(x) - \mu_i}{\log(x) - \mu'_i} \right) = 1$ , on peut réécrire

$$\lim_{x \rightarrow \infty} \left( \frac{\pi_2 \sigma_i}{\pi'_2 \sigma'_i} e^{-\frac{1}{2} \log^2(x) \left( \frac{1}{\sigma_i^2} - \frac{1}{\sigma_i'^2} \right) - \frac{1}{2} \log(x) \left( -\frac{2\mu_i}{\sigma_i^2} + \frac{2\mu'_i}{\sigma_i'^2} \right) - \frac{1}{2} \left( \frac{\mu_i^2}{\sigma_i^2} - \frac{\mu_i'^2}{\sigma_i'^2} \right)} \right) = 1.$$

Une condition nécessaire pour que cette équation soit vraie est que le terme  $\log^2(x)$  soit nul, on a donc  $-\frac{1}{2} \left( \frac{1}{\sigma_i^2} - \frac{1}{\sigma_i'^2} \right) = 0 \Rightarrow \sigma_i = \sigma'_i$ . De la même manière le terme en  $\log(x)$  doit être nul, et donc  $\mu_i = \mu'_i$ . Finalement, cela conduit à  $\lim_{x \rightarrow \infty} \left( \frac{\pi_2}{\pi'_2} \right) = 1$  et aussi  $\pi_2 = \pi'_2$ .

Il faut encore montrer l'identifiabilité des paramètres de propagation, nous utilisons pour cela la transformée de Laplace  $L(f_N) = \int_0^{+\infty} e^{st} f_N(t) dt$  of  $f_N$ . Nous avons donc

$$L(f_N) = \pi L(f_{N_p}) + (1 - \pi)L(f_{N_i})L(f_{N_p}),$$

$$L(f_N) = L(f_{N_p})(\pi + (1 - \pi)L(f_{N_i})).$$

Il a été démontré précédemment que  $\theta_i = \theta'_i$ , et donc

$$L(f_{N_p}(n; \theta_p))(\pi + (1 - \pi)L(f_{N_i})) = L(f_{N_p}(n; \theta'_p))(\pi + (1 - \pi)L(f_{N_i})),$$

ce qui implique  $L(f_{N_p}(n; \theta_p)) = L(f_{N_p}(n; \theta'_p))$ . Deux variables ayant même transformée de Laplace ont la même distribution :  $f_{N_p}(n; \theta_p) = f_{N_p}(n; \theta'_p)$ . Finalement  $\theta_p = \theta'_p$  de part l'identifiabilité des distributions de log-normale.

**2 :** Pour tout niveau de contrainte  $s$ , nous avons par définition du modèle (2.1) :

$$f_N = \pi(s) f_{N_p}(n; s, \mu_i(s), \sigma_i) + (1 - \pi(s)) f_{N_i+N_p}(n; s, \mu_i(s), \mu_p(s), \sigma_i, \sigma_p),$$

---

avec  $\pi(s) = \frac{e^{\alpha+\beta s}}{1+e^{\alpha+\beta s}}$ ,  $\mu_i(s) = a_i + b_i s$ , et  $\mu_p(s) = a_p + b_p s$ .

Pour tout  $s$ , nous avons prouvé que  $a_p + b_p s = a'_p + b'_p s$  et  $a_i + b_i s = a'_i + b'_i s$ .

Ce qui implique directement que  $a_p = a'_p$ ,  $b_p = b'_p$ ,  $a_i = a'_i$  et  $b_i = b'_i$ .

De la même manière, pour tout  $s$ , nous avons  $\frac{e^{\alpha+\beta s}}{1+e^{\alpha+\beta s}} = \frac{e^{\alpha'+\beta' s}}{1+e^{\alpha'+\beta' s}}$  et donc  $\alpha = \alpha'$  et  $\beta = \beta'$

---





---

## Annexe B

# Méthodes d'optimisations

Cette annexe présente la méthode de maximisation utilisée lors des phases M de l'algorithme EM. La méthode 'BFGS' a en effet l'avantage de ne pas nécessiter le calcul de dérivées secondes, numériquement lourd notre cas.

### B.1 Méthode de Newton-Raphson

La méthode de Newton-Raphson est un algorithme de recherche d'un maximum local d'une fonction. L'idée est la suivante : premièrement, construire une approximation quadratique de la fonction d'intérêt autour de paramètres initiaux. Ensuite, prendre comme nouveaux paramètres ceux qui maximisent l'approximation. Cette procédure est ensuite itérée jusqu'à stabilisation des paramètres. Soit  $f : \mathbb{R}^k \rightarrow \mathbb{R}$  une fonction deux fois continuellement différentiable. Soit  $x, h \in \mathbb{R}^k$ .

L'approximation au premier ordre de la formule de Taylor est donnée par :

$$f(x+h) \simeq f(x) + \nabla f(x)^T h$$

$$\text{et celle du second ordre : } f(x+h) \simeq f(x) + \nabla f(x)^T h + \frac{1}{2} h^T D^2 f(x) h,$$

où  $\nabla f(x)$  représente le gradient, et  $D^2 f(x)$  la matrice hessienne symétrique. (Respectivement, le vecteur des dérivées et la matrice des dérivées seconde.) Etant donné que nous recherchons un maximum, nous devons annuler sa dérivée. Nous pouvons donc écrire l'approximation du premier ordre de la dérivée :

$$\nabla f(x+h) = \nabla f(x) + D^2 f(x) h$$

La condition de premier ordre pour la valeur de  $h$  (notée  $\hat{h}$ ) qui maximise  $f(x+h)$  est :

$$0 = \nabla f(x) + D^2 f(x) h$$

---

Ce qui implique :  $\hat{h} = -(D^2 f(x))^{-1} \nabla f(x)$  En d'autres termes, le vecteur qui maximise l'approximation de Taylor à l'ordre 2 de  $f$  est le suivant :

$$x + \hat{h} = x - (D^2 f(x))^{-1} \nabla f(x)$$

### Algorithme de Newton Raphson

L'algorithme de Newton-Raphson est donc le suivant.

- Choisir une valeur initiale  $x_0$
- Soit  $x_i$  la valeur courante, le calcul de  $x_{i+1}$  se fait de la façon suivante :
- Le processus est itéré jusqu'à atteindre  $\|\nabla f(x_i)\| < \epsilon$  où  $\epsilon$  représente une tolérance fixée par l'utilisateur.

Une visualisation unidimensionnelle est donnée par la figure B.1

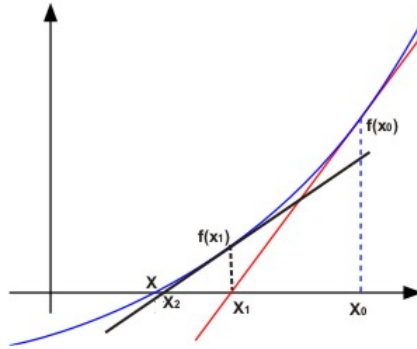


FIGURE B.1 – Schéma du principe de fonctionnement de l'algorithme de Newton-Raphson en dimension 1

*Remarque : Il est à noter que l'algorithme de Newton-Raphson ne vérifie pas les conditions de second-ordre. Ce qui signifie qu'une mauvaise initialisation peut amener sur un minimum local ou sur un point col.*

## B.2 Méthode 'BFGS'

La méthode BFGS reprend la trame de l'algorithme de Newton-Raphson, mais, au lieu de calculer à chaque étape  $D^2 f(x_i)$  elle utilise une approximation. La mise à jour de l'approximation de la matrice Hessienne est réalisée sous cette forme : Soit une fonction  $f : \mathbb{R}^k \rightarrow \mathbb{R}$  continument différentiable. Soient  $x_k$  et  $x_{k+1}$  deux itérés tel que  $d_{k-1}^T y_{k-1} > 0$  avec  $d_{k-1} = x_k - x_{k-1}$  et  $y_{k-1} = \nabla f(x_k) - \nabla f(x_{k-1})$ . Soit une matrice définie positive  $H_{k-1} \in \mathbb{R}^{k \times k}$ .

La mise à jour BFGS est donnée par :

$$H_k = H_{k-1} + \frac{y_{k-1} y_{k-1}^T}{y_{k-1}^T d_{k-1}} - \frac{H_{k-1} d_{k-1} d_{k-1}^T H_{k-1}}{d_{k-1}^T H_{k-1} d_{k-1}}$$

---

**Algorithme de quasi-Newton : BFGS**

- Soient  $x_0 \in \mathbb{R}^k$  et  $H_0 \in \mathbb{R}^{k \times k}$ . Une initialisation du vecteur et de la matrice Hessienne. En pratique on prend  $H_0 = I_{k \times k}$
  - trouver  $p_k$  tel que  $p_k = -B_k^{-1} \nabla f(x_k)$
  - Phase de recherche linéaire : choisir  $\alpha_k$  maximisant  $f(x + \alpha p_k)$
  - mise à jour :  $x_{k+1} = x_k + \alpha_k p_k$   
$$y_k = \nabla f(x_{k+1}) - \nabla f(x_k) \quad d_k = \alpha_k p_k \quad H_{k+1} = H_k + \frac{y_k y_k^T}{y_k^T d_k} - \frac{H_k d_k d_k^T H_k}{d_k^T H_k d_k}$$
-



---

## Annexe C

# Résultats supplémentaires de simulation

nombre de points	écart-type
20	$1.2 \times 10^{-3}$
50	$7 \times 10^{-4}$
100	$5 \times 10^{-4}$
200	$5 \times 10^{-4}$
400	$5 \times 10^{-4}$

TABLE C.1 – Sur 100 simulations, comportement de l'estimateur du paramètre  $b_p$  obtenu, le coefficient à l'origine de la propagation.

nombre de points	écart-type	
20	0.14	
50	0.1	
100	0.1	
200	0.09	
400	XX	0.1

TABLE C.2 – Sur 100 simulations, comportement de l'estimateur du paramètre  $\sigma_p$  obtenu, le coefficient à l'origine de la propagation. (Valeur simulée : XX)

---

nombre de points	écart-type
20	3.1
50	1.5
100	0.8
200	0.46
400	0.6

TABLE C.3 – Sur 100 simulations, comportement de l’estimateur du paramètre  $a_i$  obtenu, le coefficient à l’origine de la propagation.

nombre de points	écart-type
20	$4 \times 10^{-3}$
50	$1.4 \times 10^{-3}$
100	$1 \times 10^{-3}$
200	$7 \times 10^{-4}$
400	$1 \times 10^{-3}$

TABLE C.4 – Sur 100 simulations, comportement de l’estimateur du paramètre  $b_i$  obtenu, le coefficient à l’origine de la propagation.

nombre de points	écart-type
20	0.34
50	0.16
100	0.09
200	0.08
400	0.09

TABLE C.5 – Sur 100 simulations, comportement de l’estimateur du paramètre  $\sigma_i$  obtenu, le coefficient à l’origine de la propagation.

---