

ORSAY

N° d'ordre : 8826

UNIVERSITE PARIS-SUD
FACULTE DES SCIENCES D'ORSAY

THESE

Présentée pour obtenir

LE GRADE DE DOCTEUR EN SCIENCES
DE L'UNIVERSITE DE PARIS-SUD

Spécialité : Mathématiques

par

Marc LAVARDE

**FIABILITE DES SEMI-CONDUCTEURS, TESTS ACCELERES,
SELECTION DE MODELES DEFINIS PAR MORCEAUX ET
DETECTION DE SUR-STRESS.**

Soutenue le 7 novembre 2007 devant la commission d'examen :

M. Gilles CELEUX	(Président du jury)
M. Jean-Louis BON	(Rapporteur)
M. Fabrice GUERIN	(Rapporteur)
M. Pascal MASSART	(Directeur de thèse)
M. Patrick PAMPHILE	(Directeur de thèse)

Remerciements

Ma première pensée va à Patrick Pamphile, sans qui ce manuscrit ne serait qu'un souvenir inachevé dans un coin de mon PC. Je te remercie chaleureusement de m'avoir soutenu et motivé sans ménager ton temps. À ton contact j'ai beaucoup appris durant ces dernières années. J'ai eu beaucoup de chance et de plaisir de pouvoir travailler avec toi.

Un grand merci à Pascal Massart pour m'avoir initié aux mystères de la sélection de modèle avec forte patience et une grande disponibilité (que je regrette de ne pas avoir davantage sollicité). Merci d'avoir su me remettre sur la route et m'encourager lorsqu'il le fallait.

Je remercie Gilles Celeux, Jean-Louis Bon, Fabrice Guérin, Pascal Massart et Patrick Pamphile qui me font l'honneur de participer au jury de cette thèse. Et je remercie particulièrement Jean-Louis Bon pour ses nombreuses remarques et ces conseils et Fabrice Guérin pour avoir rapporté mon travail.

Je tiens à remercier mes anciens collègues du service Assurance Qualité d'ALTIS, merci de m'avoir fait découvrir le monde industriel et la Microélectronique. Vous m'avez montré que la solidarité habite aussi les open-spaces : j'ai une pensée amicale pour Bernt, Danielle, Gérard, Holly, Jean-Pierre, Lionel, Patrick, Régis et pour mes collègues « stagiaires » : Nathalie, Florent, Fred, et Sandra.

Merci également à l'équipe pédagogique de GEA1 de l'IUT de Sceaux pour m'avoir ouvert les portes de l'enseignement ; avec une pensée amicale pour Isabelle et Evelyne. Je remercie Gilles Celeux et l'ensemble du groupe SELECT pour leurs séminaires aux sujets aussi passionnants que variés. Merci à Marie Sauvé pour sa collaboration et pour m'avoir expliqué ses théorèmes de sélection dans les détails. Je remercie également Pierre Pansu, Isabelle Souriou et Valérie Lavigne pour leur aide dans mes démarches administratives.

Pour conclure, j'adresse un grand merci à ma famille pour leur soutien constant et sans condition. Enfin, merci à mes amis pour m'avoir motivé et supporté pendant tout ce temps.

Résumé

Cette thèse traite de l'exploitation de données accélérées et de la sélection de modèles de régression dans le domaine des semi-conducteurs. Les données recueillies à la suite d'un test accéléré sont des données de régression. L'objectif du test est d'ajuster le comportement moyen du logarithme des durées de vie à l'aide d'une fonction f , dite fonction d'accélération. Cependant les données accélérées ont parfois des comportements complexes. Afin d'adapter la modélisation à ces comportements atypiques, nous avons cherché à détecter les changements de comportement de la fonction d'accélération. Nous proposons $(\mathcal{F}_M)_{M \in \mathcal{M}_n}$ une collection de modèles de régressions définis par morceaux, pour chaque modèle candidat à l'estimation nous calculons l'estimateur des moindres carrés \hat{f}_M . Et nous sélectionnons le modèle final $\hat{f}_{\hat{M}}$ à l'aide d'un critère des moindres carrés pénalisés. L'estimateur pénalisé est une approximation optimale du modèle réel au sens où le risque de l'estimateur pénalisé $\hat{f}_{\hat{M}}$ est comparable au risque minimum parmi $(\hat{f}_M)_{M \in \mathcal{M}_n}$ l'ensemble des modèles candidats. De plus, nous disposons d'une borne de risque non asymptotique. Et nous avons cherché à limiter les hypothèses de modélisation afin de prendre en compte un grand nombre de cas pratiques.

Notre premier objectif est de modéliser la fiabilité des semi-conducteurs. Nous avons consulté la physique de la dégradation afin de définir les familles de fonctions d'accélération propres à ce domaine d'application. Nous avons listé les spécificités de production et les mécanismes de dégradation de notre contexte d'application. Puis, nous avons étudié les modèles AFT par la méthode des moindres carrés et du maximum de vraisemblance. Et nous avons discuté des cas limites, où le test révèle une accélération associée aux conditions opérationnelles et une (ou plusieurs) accélération(s) propre(s) à caractériser la fiabilité du composant sous les plus hautes conditions de stress : on parle alors de test en sur-stress ou en surmenage.

Notre second objectif est de détecter le sur-stress et de modéliser la fiabilité des semi-conducteurs. À la suite d'un test accéléré en surmenage, nous avons cherché à détecter les ruptures dans la relation d'accélération. Nous proposons de détecter les ruptures dans f la relation d'accélération et donc d'estimer f comme une fonction définie par morceaux où k le nombre d'accélération est inconnu. Nous avons utilisé un critère des moindres carrés pénalisés construit à partir de l'étude du logarithme népérien de la durée de vie et de considérations non asymptotiques. Dans un premier temps, nous avons considéré le cas Gaussien associé aux mécanismes de migration lente de matériaux. Et puis, en vue d'appliquer cette méthode à un plus grand nombre de cas, nous avons considéré le cas Gumbel associé aux défauts d'usure par chocs. Remarquons toutefois que le cadre théorique que nous avons utilisé pour établir le critère de pénalité est plus large que les simples cas Gaussien et Gumbel ; il prend en compte le cas où la transformée de Laplace de la perturbation Z est bornée au voisinage de zéro.

Contents

1	Fiabilité des Semi-conducteurs	13
1.1	Semi-conducteur	13
1.1.1	Préambule	13
1.1.2	Qu'est-ce qu'un semi-conducteur?	14
1.1.3	Du transistor à la microélectronique	14
1.1.4	Production de masse des semi-conducteurs	16
1.2	Mécanisme de fiabilité des semi-conducteurs	22
1.2.1	Détection de la défaillance	22
1.2.2	Modes de défaillance	24
1.3	Défaillance d'usure	26
1.3.1	Tests d'usure	26
1.3.2	Électromigration	27
1.3.3	Stressmigration	29
1.4	Défaillances par chocs	30
1.4.1	Test de Défaillance	30
1.4.2	Claquage d'un transistor	31
1.5	Généralités sur la fiabilité	32
1.5.1	Retour d'expérience	32
1.5.2	Standards de fiabilité	33
1.5.3	Essais de fiabilité et accélération	33
1.6	Modèles de fiabilité	34
1.6.1	Modèle et notations	34
1.6.2	Distribution Lognormale	34
1.6.3	Distribution Weibull	36
1.7	Modèle accéléré	38
1.7.1	Hypothèses et notations	38

1.7.2	Accélération classique	41
1.7.3	Modèle standard de vie accélérée	42
1.8	Modèle AFT	44
1.8.1	Notations et définitions	44
1.8.2	Estimateur de contraste du minimum	45
1.8.3	Maximum de vraisemblance	46
1.8.4	Application et Arrhénius Lognormale	47
1.8.5	Application et modèle de Weibull	49
1.9	Limites des hypothèses d'accélération	51
1.9.1	Généralités	51
1.9.2	Multiples plages d'accélération	52
2	Detection of Overstress for Lognormal ALT	55
2.1	Introduction	56
2.1.1	Generality about ALT	56
2.1.2	Introducing selection	56
2.2	Piecewise acceleration model	57
2.2.1	Lifetime distribution	57
2.2.2	Accelerated Failure Time model	58
2.2.3	Limits of the model	58
2.2.4	Piecewise model	59
2.3	Number of acceleration	59
2.3.1	Estimation of the piecewise model	59
2.3.2	Penalized least squares	60
2.3.3	Selection theorem	61
2.3.4	Form of the penalty function	62
2.3.5	Importance of the calibration	64
2.4	Theoretical calibration	65
2.4.1	Aim of the procedure	65
2.4.2	Description of the procedure	65
2.4.3	Calibration for piecewise affine regression	66
2.5	Data-driven calibration	69
2.5.1	Birgé-Massart's heuristic	70
2.5.2	Application of the heuristic	71
2.5.3	Full noisy data	74

2.6	Data-driven correction of the calibration	76
2.6.1	Aims and method of correction	76
2.6.2	Numerical application	78
2.6.3	Application on real data	80
3	Detection of Overstress for Weibull ALT	91
3.1	Introduction	91
3.2	Assumptions & Notations	94
3.3	Selection theorem	96
3.4	First method	99
3.4.1	Birgé-Massart's heuristic	100
3.4.2	Numerical application	101
3.5	Second method	104
3.5.1	Estimation of c^*	105
3.5.2	Estimation of λ^*	108
3.5.3	Second method in practice	109
3.5.4	Class-H motor insulation data	113
A	Recherche des modèles candidats	119
A.1	Partition dynamique	119
A.2	Partition CART	121
	Conclusion et perspectives	
	List of Figures	
	List of Tables	
	Index	
	Bibliography	

Introduction

Le sujet de cette thèse est issue du milieu industriel avec la collaboration de ALTIS Semiconductor (joint venture IBM-INFINEON) une fonderie de semi-conducteurs implantée à Corbeil-Essonnes. Nous avons abordé plusieurs problématiques propres à la fiabilité des composants microélectroniques. Toutefois, des problématiques similaires se retrouvent dans d'autres domaines novateurs comme les industries chimiques ou plastiques, le milieu automobile, aéronautique et bien d'autres . . .

Les tests accélérés permettent d'obtenir rapidement des informations sur la durée de vie et d'enrichir les connaissances de la physique de la défaillance. En pratique, ces tests ne sont pas seulement utilisés lors des étapes de recherche et de développement d'un produit ; ils servent également à mesurer la qualité d'un composant tout au long de sa durée d'exploitation commerciale. À court terme, les tests accélérés permettent de comparer statistiquement différents processus industriels. Et à long terme, ils permettent de comparer statistiquement différentes générations de composants. Les préoccupations de l'ingénieur sont de donner des estimations réalistes et adaptées à ce contexte pratique. Cependant comme dans tous les domaines de hautes technologies, les évolutions majeures de la Microélectronique se succèdent à un rythme élevé. Ainsi la rapidité et les profonds changements technologiques liés à cette activité rendent obsolète les modélisations utilisées lors des tests accélérés : la conséquence étant que les prévisions de fiabilité sont fausses. Et nous proposons de mettre en place une méthode statistique afin de repérer ce type de comportements.

Les données issues d'un test accéléré à environnement constant sont des données de régression. La variable réponse (observation), notée Y , est une variable réelle. Elle correspond au logarithme népérien de la durée de vie du composant testé. Et la variable explicative (covariable), notée $\mathbf{x} = (x_1, \dots, x_p)$, est un vecteur aux composantes réelles dans lequel chaque composante x_i représente une variable environnementale (stress) ou une transformation d'une (ou plusieurs) variable(s) environnementale(s). Nous supposons que les variables explicatives sont données par le plan d'expérience et à partir des connaissances des mécanismes de dégradation. Notons \mathcal{X} l'ensemble des variables environnementales, le plus souvent \mathcal{X} est \mathbb{R}^p ou un sous-ensemble de \mathbb{R}^p . Nous considérons donc le modèle Accelerated Failure Time (AFT, suivant [28, 27, 23] :

$$Y = f(x_1, \dots, x_p) + \sigma \varepsilon,$$

où $f : \mathcal{X} \rightarrow \mathbb{R}$ est une fonction linéaire inconnue qui relie les variables explicatives à la réponse, σ est un paramètre de dispersion du modèle et ε est la variable aléatoire liée

à l'erreur de modélisation (centrée et réduite). Lorsque les variables explicatives sont choisies de sorte que le phénomène d'usure est accéléré ; alors on parle de modèle accéléré. Dans ce cas, les variables explicatives sont des variables de stress. Comme f est liée à la cinétique du mécanisme d'usure, la fonction f est une fonction d'accélération. On choisit f et la loi de probabilité de ε à partir de connaissances du mécanisme de dégradation. On retrouve alors les lois usuelles utilisées en fiabilité comme la loi Normale standard ou une version standardisée de la loi de Gumbel.

Nous avons développé deux problématiques (liées) :

- Les modèles accélérés utilisés lors des études de fiabilité des semi-conducteurs ;
- La détection de rupture(s) dans la relation d'accélération par sélection de modèles.

Les techniques de sélection de modèles par estimateur pénalisé permettent de comparer l'ensemble des modèles candidats à partir d'un critère pénalisé comprenant :

- un terme de biais : lié à la précision de la prévision ;
- un terme de variance : lié à la complexité du modèles.

Dans l'article de référence de Birgé et Massart [6], on trouve un cadre théorique de la sélection de modèles dans le cas Gaussien. Le cas Gaussien est particulièrement bien adapté aux problèmes de défaillance due à l'usure. Nous avons donc mis en oeuvre une calibration de la pénalité adaptée à notre problème de fiabilité.

Cependant la défaillance par chocs, qui regroupe un large spectre de problèmes de fiabilité, n'entre pas dans le cadre Gaussien. Les travaux très récents de Sauv  [31] nous ont permis de résoudre ce problème. Les techniques et les paramètres de calibration proposés sont différents du cas Gaussien. Nous avons donc mis en oeuvre la calibration de la pénalité adaptée à ce cas non-Gaussien.

Dans le chapitre 1, nous avons étudié la fiabilité d'une famille de composants de haute technologie : les semi-conducteurs. Dans cette partie, nous décrivons les spécificités des semi-conducteurs en tant que matériau doté de nombreuses propriétés physiques et en tant que composant élémentaire des composants microélectroniques. Puis nous parlons des mécanismes de dégradation. Ensuite nous abordons le modèle AFT et l'estimation de la fiabilité à partir de données issues de tests accélérés. Nous introduisons le problème des données accélérées en sur-stress où un phénomène de surmenage fait apparaître un artefact dans les données accélérées : on observe alors une (ou plusieurs) rupture(s) dans la relation d'accélération.

Dans les chapitres 2 et 3, nous avons traité le problème de la détection de ruptures dans la fonction d'accélération. Dans le chapitre 2, nous avons proposé une méthode de détection adaptée au cas où la distribution de probabilité de la durée de vie est Lognormale. Nous avons implémenté un critère des moindres carrés pénalisés, et nous avons utilisé une fonction de pénalité construite à partir du théorème de sélection de Birgé et Massart [6]. Nous avons appliqué plusieurs méthodes de calibrations de la fonction de pénalité lorsque σ est donné : la méthode de calibration de Birgé et

Rozenholc [8], la méthode de calibration de Birgé et Massart [6] et sa variante proposée par Lebarbier [19] afin d'intégrer une estimation du paramètre de dispersion σ . Et nous avons comparé ces méthodes à AIC et BIC dans le cadre de la détection des plages d'accélération d'un test accéléré.

Dans le chapitre 3, nous avons utilisé une méthode de détection adaptée au cas où la variable aléatoire Y suit la première loi des valeurs extrêmes : la loi de Gumbel. Nous avons proposé un critère des moindres carrés pénalisés et nous avons utilisé une fonction de pénalité construite à partir du théorème de sélection de Sauvé [31, 32] qui généralise le résultat de Birgé et Massart [6] dans un cadre non-Gaussien (en particulier, il est valable dans le cas où la distribution de Y est Gumbel). Puis nous avons mis en place la sélection de modèles dans le cadre de la détection de rupture de l'accélération lors d'un test accéléré. Et nous avons implémenté la méthode de calibration *data-driven* de Birgé et Massart [5, 6]. En nous inspirant de la méthode de calibration de Lebarbier [19], nous avons proposé une méthode de calibration adaptée à la nouvelle forme de la fonction de pénalité et basée sur une étude de données de pur bruit et l'heuristique de Birgé-Massart. Enfin nous avons comparé les performances des méthodes avec les critères classiques (AIC et BIC) et nous avons étudié une correction adaptative de la fonction de pénalité BIC en nous inspirant de l'heuristique de Birgé-Massart.

Fiabilité des Semi-conducteurs

Un semi-conducteur est un matériaux qui n'est ni conducteur, ni isolant. Par abus de langage, on parle souvent de semi-conducteur alors qu'on désigne un circuit microélectronique intégré sur un substrat semi-conducteur ultra pur. Les études de fiabilité des semi-conducteurs rencontrent deux difficultés majeures :

- La complexité des composants : constitués de multiples matériaux et la miniaturisation dont les limites sont repoussées d'une génération à l'autre ;
- Les technologies matures et innovantes : des durées de vies à long terme.

Dans le chapitre 1, nous débutons donc notre étude en exposant les propriétés des semi-conducteurs, leur rôle dans la miniaturisation électronique et nous parlerons de la fabrication d'un circuit intégré sur un substrat de semi-conducteur ultra pur (le silicium). Ensuite, nous présentons les principaux modes de défaillances des semi-conducteurs, sous l'angle de la physique de la dégradation et avec un souci de vulgarisation. Nous avons utilisé ces modes de défaillances afin de justifier les hypothèses des modèles de fiabilité. Puis, nous abordons le REX et les standards industriels, avant de parler en détails des tests accélérés et des modèles associés à ce genre de test de fiabilité.

Comme nous l'avons rappelé précédemment, les tests accélérés sont des tests réalisés sous différents environnements propres à stresser le composant. Par exemple, de telles circonstances sont souvent réunies suite à une augmentation de la température, de la pression atmosphérique, du taux d'humidité, de l'intensité ou du voltage d'un circuit

électrique, ... Dans ce genre de test, on recherche à accroître les contraintes environnementales afin d'optimiser le temps du test et donc on cherche à accélérer le processus de vieillissement du composant. Par ailleurs, le protocole expérimental des tests accélérés introduit un effet de bord intéressant ; comme l'étude des données issues de tests accélérés permet d'estimer les caractéristiques de fiabilité de l'unité de test sous différentes conditions environnementales, on dispose alors d'informations sur la fiabilité du composant pour une certaine gamme de stress. Cet apport peut sembler dérisoire lorsque l'on s'intéresse au système électronique d'un téléphone portable, mais il devient indispensable lorsque l'on s'intéresse au système électronique embarqué dans un avion. De plus, ces données enrichissent le REX et elles permettent de renouveler les standards industriels.

Nous discutons dans ce document des modèles accélérés en supposant que les conditions environnementales sont fixées et donc les stress sont des variables explicatives. Cependant, il existe d'autres tests accélérés, comme le strep-stress où les environnements de stress augmentent avec le temps : soit par paliers, soit de manière continue. On suppose alors que les stress sont des variables aléatoires. Nous présentons les fonctions d'accélération classiques. Ensuite, nous avons rappelé les méthodes d'estimations appliquées aux modèles accélérés dans un cadre général, nous avons utilisé la méthode des moindres carrés et le maximum de vraisemblance.

Enfin, nous discutons de plusieurs cas limites souvent rencontrés à la suite d'une innovation technologique (nouveaux matériaux, nouveaux processus d'intégration ...). Dans de tels cas, les données issues du test accéléré peuvent être fortement atypiques. En règle générale, on rencontre deux cas particuliers, soit on observe un (ou plusieurs) mode(s) de défaillance(s) alternatif(s), soit on observe que les évolutions techniques modifient la cinétique de la dégradation pour les hauts niveaux de stress et on observe alors plusieurs plages d'accélération. Dans le premier cas particulier, on parle de problème de risques compétitifs (voir Nelson [27]). Ce phénomène est bien connu et il n'est pas propre aux tests accélérés. Le second cas particulier est spécifique aux tests accélérés. Dans ce cas, le test accéléré induit un phénomène de surcharge, également connu sous le terme (anglicisme) de sur-stress. On voit donc apparaître des ruptures dans la relation d'accélération et chaque plage (définies par lesdites ruptures) traduit un comportement propre à révéler une accélération définie pour un sous-ensemble de conditions environnementales données. Si on ne prend pas en compte de ce genre de phénomènes dans le modèle d'accélération, alors les estimations des caractéristiques de fiabilité sont fortement erronées. Et souvent, on sous-estime la qualité des unités tests. Dans les chapitres 2 et 3, nous poursuivons l'analyse de ce phénomène de sur-stress des tests accélérés et nous mettons en place des méthodes statistiques de détection des plages d'accélération.

Introduction à la sélection de modèles

Le problème de données accélérées en sur-stress mène à considérer plusieurs plages d'accélération. Notons J_1, J_2, \dots, J_k les k plages d'accélération liées à un tel test accéléré. Soit $((Y_1, \mathbf{x}_1), \dots, (Y_n, \mathbf{x}_n))$ un n -échantillon *i.i.d.* issu de données accélérées

en sur-stress, alors le modèle d'accélération est défini par morceaux est donc :

$$Y_i = \sum_{q=1}^k f_q(\mathbf{x}_i) 1_{(\mathbf{x}_i \in J_q)} + \sigma \varepsilon_i$$

où $f_q : \mathcal{X} \rightarrow \mathbb{R}$ est la fonction d'accélération associée à la plage d'accélération J_q .

A priori, k le nombre de plages d'accélération est inconnu, tout comme les plages d'accélération J_1, J_2, \dots, J_k . Pour des modèles de même complexité, le critère des moindres carrés permet d'estimer f_1, f_2, \dots, f_k et J_1, J_2, \dots, J_k . On dispose alors le meilleur candidat des modèles à k plages d'accélération. Cependant, lorsque l'on souhaite sélectionner un modèle parmi une famille de modèles de complexité inconnue, le critère des moindres carrés ne suffit pas. On a souvent recours à un critère pénalisé. Supposons que le nombre maximal de plages d'accélération est N_n , et définissons la fonction de pénalité $pen : \mathbb{R}^p \rightarrow \mathbb{R}^+$ comme étant une fonction positive et décroissante adaptée aux hypothèses d'accélération. Alors, on peut calculer \hat{k} l'estimateur des moindres carrés pénalisés comme suit :

$$\hat{k} = \arg \min_{k=1, \dots, N_n} \left[\frac{1}{n} \sum_{i=1}^n (Y_i - \hat{f}_k(\mathbf{x}_i))^2 + pen(k) \right].$$

Les fonctions de pénalité classiques, telles que AIC ou BIC, ont été construites à partir de considérations asymptotiques. Et lorsque n est petit ou lorsque N_n le nombre de modèles considérés est très grand, alors dans de nombreux cas, on constate que ces critères pénalisés surestiment (en terme de dimension) le modèle. Voilà pourquoi nous avons recours à une correction de ces pénalités classiques.

Suite aux travaux de Talagrand sur les inégalités de concentration, Baron, Birgé et Massart [5] ont proposé une technique qui permet de construire une fonction de pénalité en se basant sur des considérations non asymptotiques. Les auteurs ont aussi proposé des méthodes d'ajustement de la pénalité à partir des données afin de construire un critère pénalisé adaptatif. Nous avons appliqué ce type de méthode afin de détecter le sur-stress dans les deux principaux contextes de fiabilité : le cas où la distribution de probabilité de Y est Normale ou de Gumbel.

Sélection de Modèles Accélérés Lognormaux par morceaux et Détection du Sur-Stress

Considérons que les données sont issues d'un test accéléré et que la loi de probabilité des durées de vies est Lognormale. Donc la loi de probabilité de l'erreur ε est Normale et à σ connu. Ce contexte est très exactement celui dans lequel Birgé et Massart [6, 7, 21] ont établi une série de résultats théoriques et pratiques concernant la construction d'un critère pénalisé non asymptotique. Nous avons supposé que les modèles candidats sont sélectionnés à partir de l'ensemble des fonctions accélérées à k plages d'accélération avec $1 \leq k \leq N_n$. Dans ce cas, la forme générale de la fonction de pénalisé, notée pen_n , est la suivante :

$$pen_n(k) = \sigma^2 \frac{r k}{n} \left[c_1 \log \left(\frac{N_n}{k} \right) + c_2 \right],$$

où r désigne le nombre de paramètres des fonctions d'accélération f_q et le couple (c_1, c_2) est composé de deux constantes de calibrations. Par construction, le terme $\sigma^2 \frac{rk}{n} c_1 \log\left(\frac{N_n}{k}\right)$ est un terme propre à corriger un biais lié à la complexité des modèles de même dimension et le terme $\sigma^2 \frac{rk}{n} c_2$ propre à corriger un biais lié à la loi de probabilité des erreurs.

Ensuite, il reste à calibrer les constantes théoriques (c_1, c_2) définissant la fonction de pénalité. Dans un premier temps, nous avons utilisé la procédure de Birgé et Rozenholc [8] afin d'estimer les constantes de calibration au plus près du théorème de sélection, c'est à dire en étudiant le rapport de risque de l'estimateur pénalisé et le risque oracle. Nous avons utilisé une méthode heuristique proposée par Birgé et Massart [6] et nous avons testé les variables initiales de cette méthode adaptative. En pratique, nous ne disposons pas toujours de la valeur exacte de σ , et d'une manière générale, il est recommandé de corriger l'échelle de valeur de la fonction de pénalité en fonction des données. Notons g_n la fonction de pénalité théorique est calibrée à l'aide d'une étude préalable, la pénalité finale est :

$$\begin{aligned} pen_n &= \alpha g_n \\ \text{et} \\ g_n(k) &= \frac{rk}{n} \left[c_1^* \log\left(\frac{n}{k}\right) + c_2^* \right] \end{aligned}$$

où le scalaire α est un terme correctif adapté à l'échantillon (Y, \mathbf{x}) et où le couple (c_1^*, c_2^*) désigne les constantes de calibration préalablement calibrées. Et enfin, nous avons appliqué cette méthode de sélection dans un cas simulé de sur stress Arrhénius Lognormale (avec accélérations) et nous avons comparé l'estimateur des moindres carrés pénalisés avec les méthodes classiques :

- La pénalité du critère AIC est la suivante : $pen_{\text{AIC}}(k) = 2\sigma^2 \frac{rk}{n}$
- La pénalité du critère BIC est la suivante : $pen_{\text{BIC}}(k) = \log(n)\sigma^2 \frac{rk}{n}$ et
- La pénalité du critère AICc est la suivante : $pen_{\text{AICc}}(k) = \sigma^2 \frac{(n+r.k_M)}{n(n-r.k_M-2)}$.

Nous avons proposé une version corrigée de BIC en utilisant l'heuristique de Birgé et Massart avec $g_n(k) = \log(n)$.

Sélection de Modèles Accélérés Weibull par morceaux et Détection du Sur-Stress

Dans cas ALT où les durées de vies suivent une loi de probabilité Weibull, la loi de probabilité de Y est de Gumbel. Ce contexte est un cas particulier de celui dans lequel Sauvé [31] a établi une version non Gaussienne du résultat de Birgé et Massart [6]. En fait, on retrouve des résultats théoriques comparables à ceux du cas Gaussien lorsque l'on réalise une recherche exhaustive parmi les modèles à k plages d'accélération avec

$1 \leq k \leq N_n$. Dans ce cas, la forme générale de la fonction de pénalité, notée pen_n , est la suivante :

$$pen_n(k) = (\tau^2 + R.b) \frac{r k}{n} \left[c_1 \log \left(\frac{N_n}{k} \right) + c_2 \right],$$

où le couple (τ, b) est caractérisé par σ et la distribution des erreurs (sous-gaussiennes lorsque $\tau = \sigma$ et $b=0$), R est une borne supérieure de l'accélération f , r désigne le nombre de paramètres des fonctions d'accélération f_q et le couple (c_1, c_2) est composé de deux constantes de calibrations. Par construction, le terme $\sigma^2 \frac{r k}{n} c_1 \log \left(\frac{N_n}{k} \right)$ est un terme propre à corriger un biais lié à la complexité des modèles de même dimension et le terme $\sigma^2 \frac{r k}{n} c_2$ est propre à corriger un biais lié à la loi de probabilité des erreurs.

Ensuite, il reste à calibrer les constantes théoriques (c_1, c_2) définissant la fonction de pénalité. Ici, nous n'avons pas utilisé la procédure de Birgé et Rozenholc [8] qui est complexe en termes algorithmiques. Nous avons appliqué l'heuristique de Birgé et Massart [6] sur des données de pur bruit afin d'estimer les constantes théoriques (c_1, c_2) . Puis, nous avons utilisé une méthode heuristique proposée par Birgé et Massart [6] ; nous avons testé les variables initiales de cette méthode adaptative. En pratique, on ne dispose pas toujours de la valeur exacte de σ et d'une manière générale, il est recommandé de corriger l'échelle de valeur de la fonction de pénalité en fonction des données. Notons g_n la fonction de pénalité théorique est calibrée à l'aide d'une étude préliminaire, la pénalité finale est :

$$\begin{aligned} pen_n &= \alpha g_n \\ &\text{et} \\ g_n(k) &= \frac{r k}{n} \left[c_1^* \log \left(\frac{n}{k} \right) + c_2^* \right] \end{aligned}$$

où le scalaire α est un terme correctif adapté à l'échantillon (Y, \mathbf{x}) et où le couple (c_1^*, c_2^*) désigne les constantes de calibration préalablement calibrées. Enfin, nous avons appliqué la méthode de sélection dans un cas simulé de sur stress Arrhénius Weibull (avec 3 accélérations) et nous avons comparé l'estimateur des moindres carrés pénalisés avec les méthodes classiques AIC, BIC et AICc. Nous avons proposé une version corrigée de BIC en utilisant l'heuristique de Birgé et Massart.

Chapter 1

Fiabilité des Semi-conducteurs

1.1 Semi-conducteur

Cette partie est largement enrichie d'informations du livre d'Henri Lilen ([20], 2004) concernant l'histoire des semi-conducteurs et également de l'expérience que j'ai acquise au sein du service Qualité d'ALTIS Semiconductors. Ce chapitre compte un grand nombre d'anglicismes propres à la microélectronique, qu'il est difficile de traduire en français tout en conservant la pertinence technique.

1.1.1 Préambule

Depuis la révolution informatique, le marché électronique est tel qu'il apparaît avantageux, à la fois d'un point de vue économique, et d'un point de vue technique, de regrouper sur un même support des circuits intégrés. En effet, ce regroupement de fonctionnalités permet de concentrer les composants produits et les efforts de rendement sur une seule et même ligne de production. Il permet également d'améliorer les performances et la qualité du produit. Par ailleurs, à mesure que la miniaturisation progresse, on voit apparaître de nouveaux champs d'application d'appareillages portables tels que les ordinateurs, les appareils GSM, les puces microélectroniques, l'électronique embarquée en automobile, les équipements médicaux, ... etc.

Les progrès dans la connaissance et dans la production des matériaux semi-conducteurs sont les points clés des avancées technologiques de la miniaturisation. À cet égard, il apparaît clairement que les semi-conducteurs sont les éléments fondamentaux, au sens propre comme au sens figuré, des sciences microélectroniques et des nanotechnologies.

L'objectif de cette partie est de présenter le contexte de notre travail d'un point de vue technologique. Nous discuterons dans un premier temps des matériaux semi-conducteurs. Ensuite nous étudierons leur rôle fondamental dans la miniaturisation électronique par le biais du transistor, puis des circuits intégrés. Nous évoquerons alors le passage de la microélectronique vers la production de masse. Enfin, nous discuterons des procédés d'intégration qui mettent en évidence une double complexité. D'une part, il y a une complexité liée aux multiples matériaux qui composent les structures. Et d'autre part, nous verrons que les contraintes de réduction de la dimension et

l'état d'avancement dans la ligne de production augmentent la complexité des procédés d'intégration.

1.1.2 Qu'est-ce qu'un semi-conducteur?

Comme l'indique leur nom, un semi-conducteur est un matériau qui n'est ni conducteur, ni isolant. Il s'agit d'un corps cristallin simple ou composé, dans un état stable. Cependant, cet état peut basculer vers un état conducteur sous l'impulsion d'un champ électromagnétique dont le potentiel dépend de variables environnementales, comme la température. Le tableau 1.1 donne une liste de quelques semi-conducteurs avec leur nom, leur symbole, la largeur de bande interdite à 300°K qui représente le seuil d'absorption d'énergie avant changement d'état et la température de fusion.

Semi-conducteur	Symbole	Largeur de bande interdite à 300°K (en eV)	Température de fusion (en °C)
Silicium	Si	1,11	1420
Germanium	Ge	0,67	937
Phosphure de gallium	GaP	2,24	1350
Arséniure de gallium	GaAs	1,43	1237
Antimoniure d'indium	InSb	0,18	525
Sulfure de zinc	ZnS	3,67	1650
Sulfure de cadmium	CdS	2,42	1750
Tellurure de cadmium	CdTe	1,61	1098
Sulfure de plomb	PbS	0,39	1103

Table 1.1: Liste de quelques semi-conducteurs.

Dans un premier temps, les semi-conducteurs étaient utilisés dans la fabrication de résistances ; on utilisait en particulier le sélénium, dont les propriétés électromagnétiques étaient bien connues suite aux travaux de Hittorf en 1851. Par la suite, on exploitait les propriétés de détection des semi-conducteurs dans la radio diffusion. Des composés de galène servaient alors à fabriquer les fameux postes à galène. Ensuite, on remplaça la galène par le silicium qui s'avérait être à la fois plus performant en termes de propriétés physiques et de rendement.

1.1.3 Du transistor à la microélectronique

Grâce aux connaissances sur l'électromagnétique des matériaux, de nouveaux composants actifs (comme le transistor, le condensateur, ...) apparaissent au lendemain de la seconde guerre mondiale. En couplant ces avancées avec les progrès techniques liés à la fabrication des composants électroniques, il était alors possible de regrouper un grand nombre de composants en un seul et unique bloc : on parle alors de circuit intégré. À ce titre, les progrès réalisés dans la fabrication des transistors illustrent parfaitement le rôle des semi-conducteurs dans l'électronique. De proche en proche, la miniaturisation des transistors met en perspective le passage vers la microélectronique.

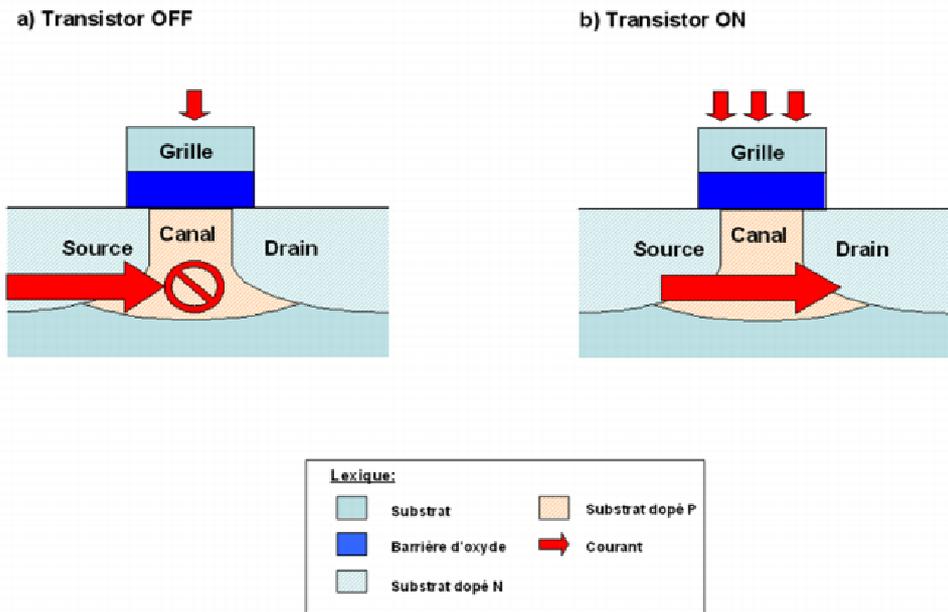


Figure 1.1: Schéma du transistor MOS (pour Metal Oxide Semiconductor) en fonction sur les schémas ; en a/ le courant est stoppé comme la tension sur la grille est nulle et en b/ on applique du courant à la grille et le transistor laisse passer le courant.

Le principe du transistor a été inventé par Julius Edgar Lilienfeld en 1930. Le nom transistor vient de la contraction de *transfer* et de *resistor*. La fonction d'un transistor est de contrôler le passage du courant à l'aide d'un potentiel électrostatique ; il s'agit d'une sorte d'interrupteur électronique. Le montage d'un transistor est illustré par la figure 1.1 montrant un transistor moderne (de type MOS à effet de champs). Le montage d'un transistor s'organise autour d'une région semi-conductrice située entre deux contacts métalliques, nommés respectivement la source et le drain. La zone semi-conductrice est une zone tampon communément appelée le canal et les propriétés électriques du composant sont directement héritées des propriétés conductrices du semi-conducteur. Pour finir le montage, on place une électrode de commande, nommée grille, au dessus du canal. Selon le potentiel de l'effet de champ appliqué sur la grille : soit le canal laisse passer le courant, soit il lui interdit le passage.

La production des premiers transistors bipolaires (1947) et des circuits intégrés (1959) marquent les débuts de la miniaturisation. Le concept du circuit intégré est dû à Jack Kulby. Il s'agit d'un composant construit d'un seul bloc solide sur la base d'un substrat en matériau semi-conducteur. L'innovation majeure du circuit intégré réside dans le fait que Kulby a intégré ses composants sur une seule plaque de germanium et qu'ensuite il a relié son circuit à l'aide de connections en or. Suite à la découverte du circuit intégré, les chercheurs de la Silicon Valley vont poser les bases de la microélectronique. En quelques années, le monde des semi-conducteurs se dirige vers une production de masse par le biais de découvertes telles que le procédé d'épitanie

(1960), la passivation, le procédé planar (1960) et le transistor MOS (1962).

1.1.4 Production de masse des semi-conducteurs

Les circuits intégrés de Kulby évoluent très vite à mesure que l'on découvre de nouveaux procédés d'intégration, c'est-à-dire des procédés liés à l'intégration de composants en un seul bloc sur une plaque de semi-conducteur ultra pur. Dans un souci de vulgarisation, on peut dire que les circuits intégrés de Kulby passent d'une architecture en 2D à une architecture en 3D grâce au procédé *planar* de Noyce (1959). Ce procédé utilise des jonctions polarisées positivement (P) et négativement (N) afin d'intégrer côte à côte des transistors et il protège le circuit des agressions extérieures par le biais d'une couche isolante nommée couche de passivation. Nous avons illustré ce procédé en représentant sur la figure 1.2 un schéma de coupe d'un circuit construit en utilisant le procédé planar. Cette idée est à la base des circuits intégrés modernes où on distingue à la base du circuit les éléments logiques séparés deux à deux par des jonctions NP.

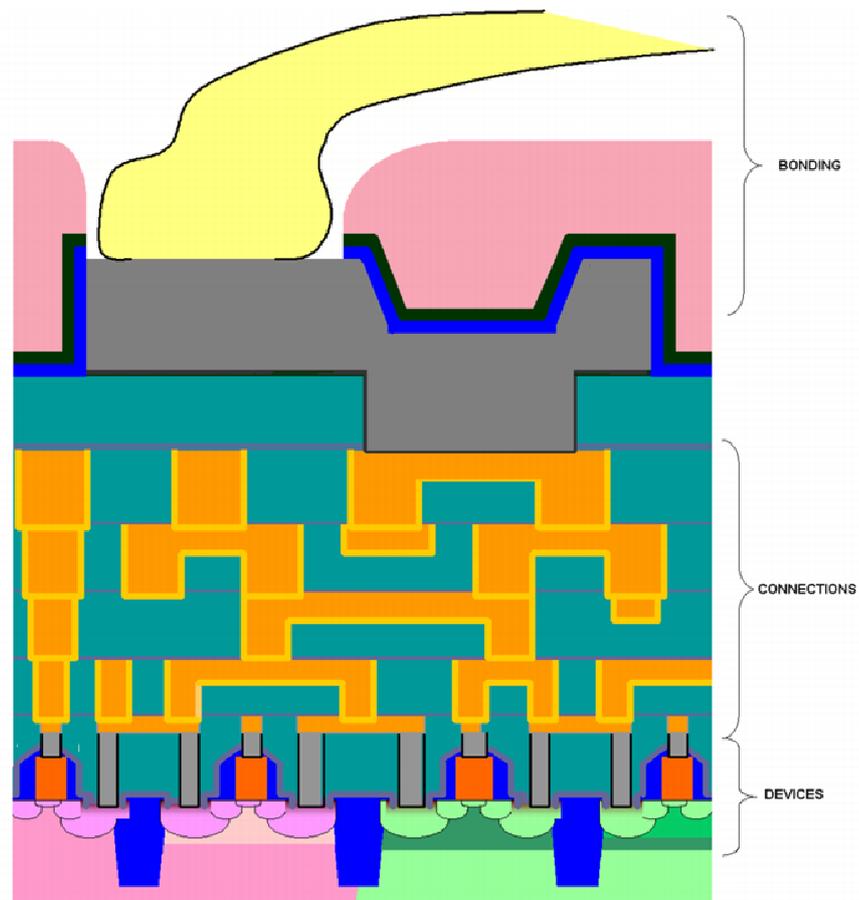


Figure 1.2: Schéma de coupe d'un circuit intégré moderne à plusieurs niveaux de connexions métalliques.

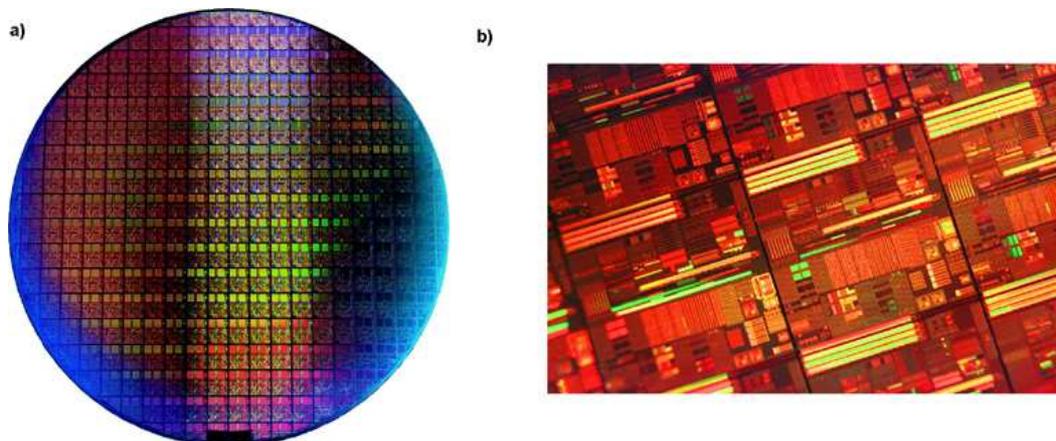
En somme ces procédés d'intégration permettent de créer en cinq phases un circuit intégré :

1. Production du substrat en semi-conducteur (aujourd'hui en silicium) ;
2. Intégrer les éléments logiques sur la base de silicium ;
3. Isoler et encapsuler les éléments actifs ;
4. Intégrer les connexions sur les éléments actifs ;
5. Intégrer une couche de passivation et les connexions externes.

Du point de vue de l'intégration, les circuits microélectroniques sont constitués d'une accumulation de couches isolantes, conductrices et semi-conductrices déposées et organisées sur une tranche de semi-conducteur (du silicium) afin de construire les structures spécialisées du composant. C'est pourquoi, on parle abusivement de semi-conducteurs tant lorsqu'on parle d'un matériau semi-conducteur, que lorsqu'on désigne un circuit intégré sur la base d'un semi-conducteur. On voit sur le schéma 1.2, un plan de coupe d'un circuit intégré. À la base on distingue des transistors MOS qui sont les structures actives du circuit représenté, puis on distingue plusieurs niveaux de connexions et enfin la partie finale constituée d'une couche de passivation et d'une connexion externe.

Maintenant, décrivons en quelques nombres clés les circuits intégrés modernes. Aujourd'hui, on intègre entre 400 et 3000 puces sur une tranche de silicium de 300 mm de diamètre et 650 μm . Les circuits intégrés sont fabriqués à la suite de 400 à 600 processus d'intégration qui s'étalent sur 60 à 90 jours. On voit sur la figure 1.3 une tranche de silicium sur laquelle on a intégré des puces microélectroniques (en vue large et rapprochée).

Figure 1.3: a/ Puce microélectronique intégrée sur une plaque de silicium; et b/ Zoom avant sur la surface de la puce.



D'une manière générale, on constate les progrès de la miniaturisation au regard du nombre de transistors constituant le circuit intégré. Ce chiffre est significatif, car le plus souvent la taille d'une puce reste du même ordre de grandeur, à savoir la taille d'un ongle. De plus, le nombre de transistors constituant un circuit intégré permet de mesurer la concentration des éléments actifs tant pour les puces mémoires que pour les puces logiques. Donc, il apparaît raisonnable de considérer le nombre de transistor par circuit comme un indicateur de complexité du composant microélectronique. Un autre indicateur du degré d'intégration est le trait qui désigne la taille du plus petit motif intégré au circuit. En règle générale, le trait est la largeur de la plus petite grille du transistor intégrée sur le circuit. Les avancées rapides de la miniaturisation sont illustrées par le tableau 1.2 dressant une liste non exhaustive de processeurs de la marque INTEL.

Processeur	Date	Nombre de transistors	Traits (en μm)
4004	1971	3.500	10
8080	1974	6.000	6
8086	1977	29.000	3
386	1985	275.000	1,5 à 1
486	1989	1.200.000	1
Pentium	1993	3.100.000	0,8
Pentium II	1997	7.500.000	0,35
Pentium III	1998	9.500.000	0,25 à 0,18
Pentium IV	2000	42.000.000	0,18 à 0,13

Table 1.2: Évolution des semi-conducteurs par une liste de quelques processeurs INTEL.

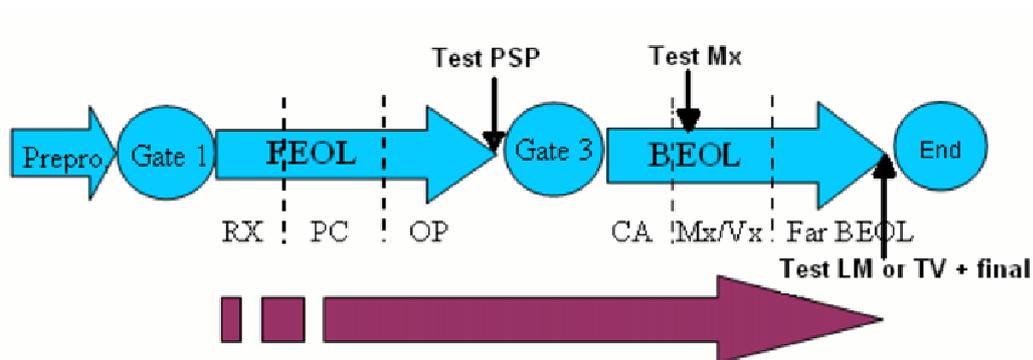


Figure 1.4: Schéma d'organisation de la ligne de production.

Les lignes de production de semi-conducteurs sont composées de deux grandes étapes de manufacture 1.4:

- *Front End Of Line* (FEOL): intégration les éléments actifs du circuit (transistor, condensateurs, ...);
- *Back End Of Line* (BEOL): intégration les éléments passifs du circuit (connexions, isolants, ...).

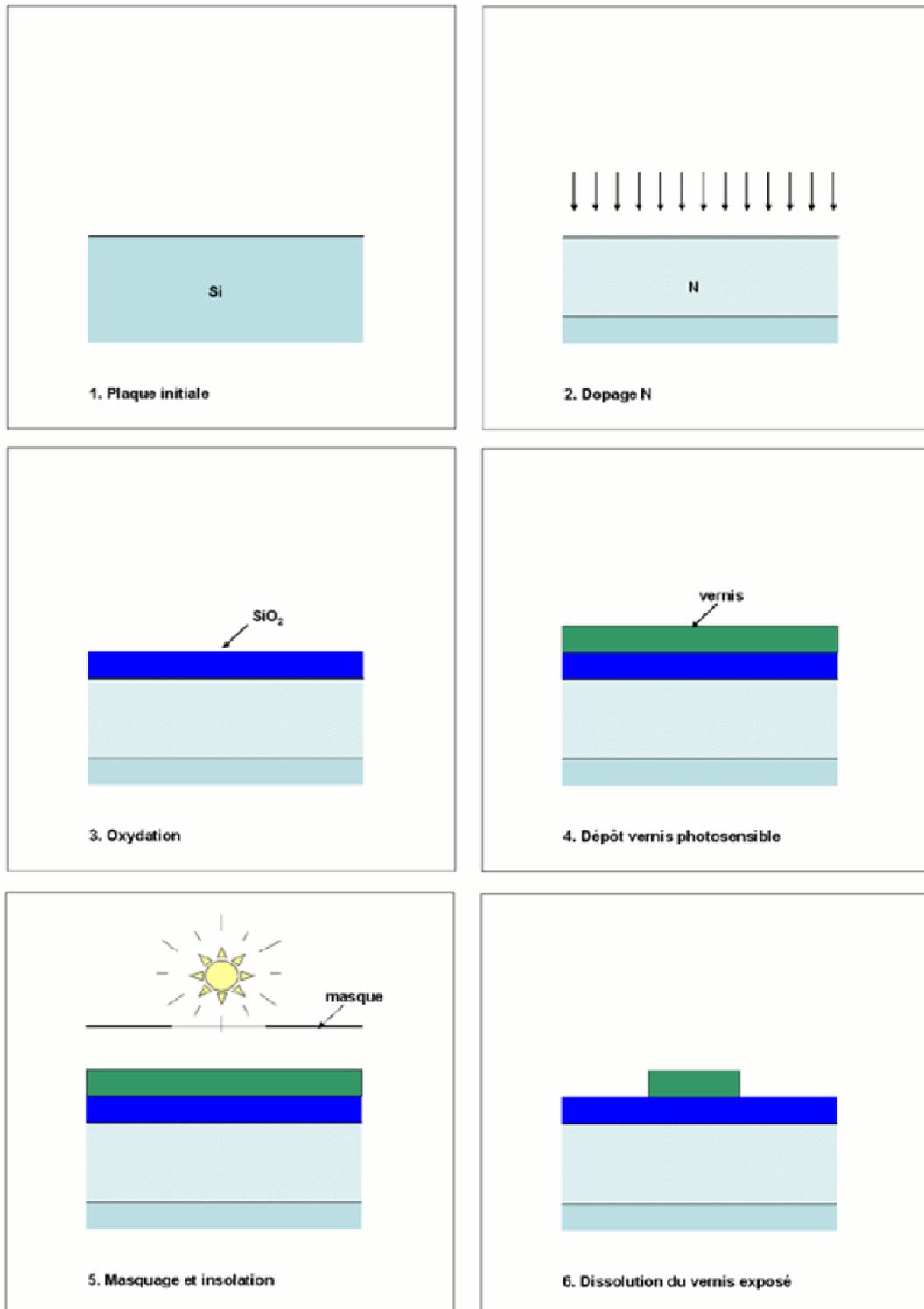
Au regard du schéma 1.4 qui représente la ligne, on constate que le FEOL se décompose en trois sous étapes : le RX où on crée les isolations entre composants actifs, le PC où on intègre les éléments actifs et enfin le OP où on isole les composants. Le BEOL se décompose également en trois sous étapes : le CA où on intègre les contacts des éléments actifs, puis on retrouve x étapes Mx/Vx où on implante un niveau x de métal et un niveau x de jonctions entre deux niveaux, et enfin le Far BEOL où on intègre la couche de passivation et les connexions externes. Chaque étape de production est jalonnée de toute une série de tests métrologiques dont le but est d'assurer un rendement constant et de garantir la qualité du composant. L'ensemble de la ligne de production est systématiquement placé en salle blanche. En effet, comme les procédés d'intégration sont très précis, il faut éviter qu'il y ait le moindre élément contaminant. De plus, les procédés d'intégration requièrent un contrôle total des conditions environnementales (température, pression, ...).

Au final, on retrouve une puce microélectronique comme sur la figure 1.2. À la vue de ce schéma représentant l'intégration d'un transistor MOS, on constate que les circuits intégrés sont construits à la suite d'un empilement structuré de matériaux conducteurs, isolants et semi-conducteurs. Un certain nombre de procédés élémentaires permettent d'intégrer les composants actifs et les composants passifs nécessaires à la fabrication de la puce microélectronique.

La figure 1.5 nous montre les procédés (vulgarisés) nécessaires à l'intégration d'un transistor MOS comme sur la figure 1.1.

Voici quelques uns des procédés élémentaires d'intégration les plus utilisés :

- *Dépôt* : ce procédé permet de déposer physiquement ou chimiquement de la matière ;
- *Dopage* : ce procédé consiste à implanter des impuretés dans la structure cristalline d'une surface cible afin de modifier ses propriétés électriques ;
- *Épitaxie* : ce procédé consiste à porter à haute température, puis à la faire décroître par plateau afin de modifier la structure cristalline. L'épitaxie est également utilisée pour modifier la structure des interconnexions et des jonctions métalliques ;
- *Gravure* : ce traitement chimique (acide) ou physique (par projection de plasma) permet de dégager des zones solides ;



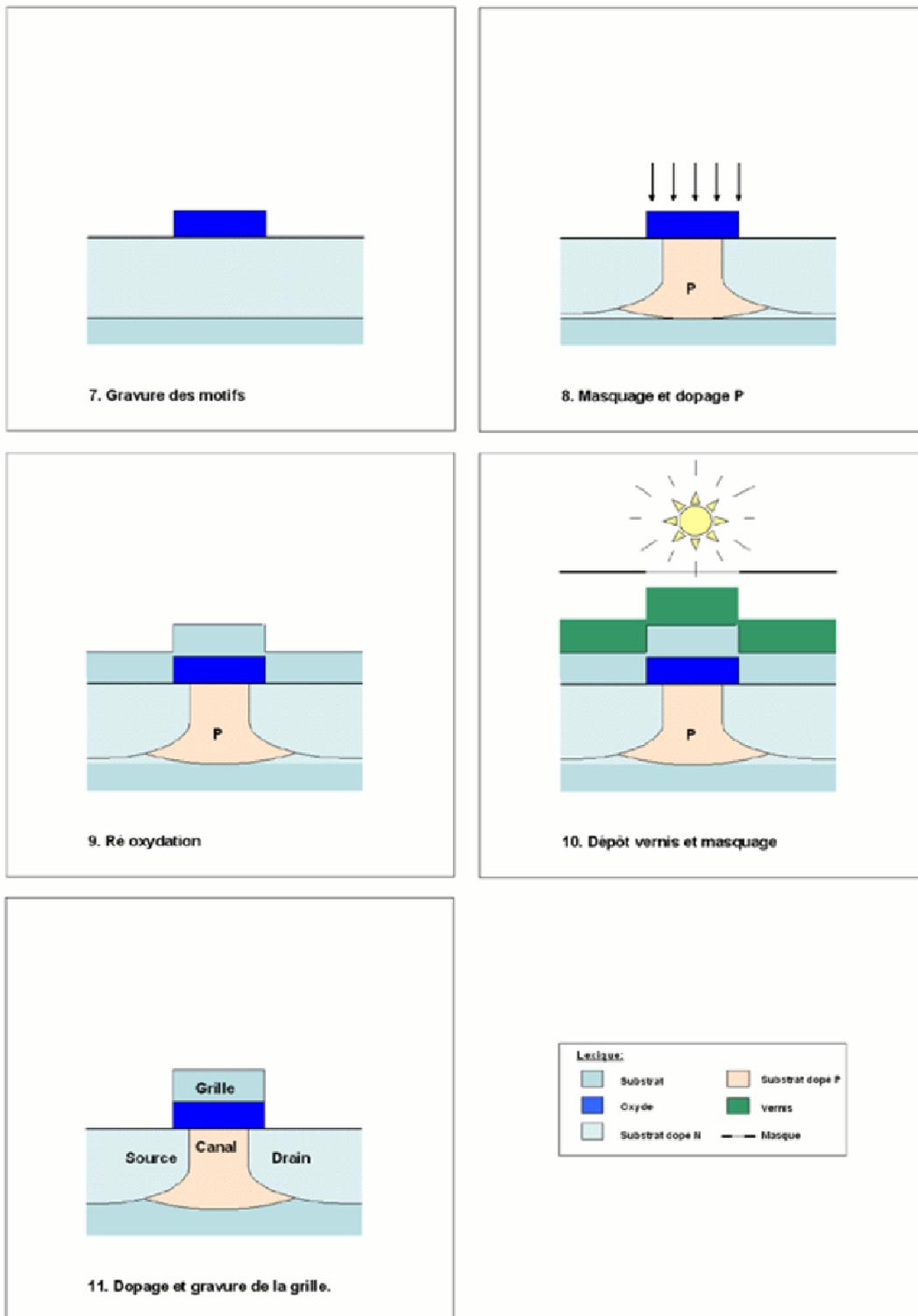


Figure 1.5: Schéma simplifié des processus d'intégration nécessaires à la fabrication d'un transistor.

- *Photolithographie*: cette opération de masquage, par des matières photosensibles, combinée à la projection d'UV, permet de tracer des motifs sur une surface donnée en vue de gravure, de dépôt ou de dopage ;
- *Oxydation* : ce processus est un traitement chimique oxydant souvent effectué à chaud ;
- *Passivation* : ce procédé cité plus haut consiste à emballer une structure dans une couche d'oxyde afin de la protéger de tout contaminant ; ... etc

Remarque 1.1.1. Il existe souvent une grande variété de procédés d'intégration pour une même opération. En premier lieu, ce fait se comprend aisément au regard du grand nombre de matériaux à intégrer au circuit. D'autre part, cette variété reflète la volonté de recherche afin d'optimiser les rendements de production. Enfin, comme les procédés s'enchaînent, il est nécessaire de prendre la mesure de l'agressivité (en terme de contraintes mécaniques et thermodynamiques) des procédés sur les structures préalablement intégrées, en vue d'organiser une suite de procédés d'intégration optimale dans le sens où il réalise un compromis entre temps, coût et qualité.

1.2 Mécanisme de fiabilité des semi-conducteurs

1.2.1 Détection de la défaillance

À présent que nous avons en tête le circuit intégré dans sa complexité, nous pouvons nous intéresser aux mécanismes de défaillance des semi-conducteurs.

Tout d'abord, regardons comment définir la défaillance d'un semi-conducteur et de circuits intégrés. Selon l'International Electrotechnical Commission, la *défaillance* est définie dès lors qu'il y a «*cessation de l'aptitude d'une entité à accomplir une fonction requise*». Dans l'ingénierie, la défaillance consiste en un changement d'état : d'un état de fonctionnement usuel vers un état dégradé ou de non fonctionnement. Ainsi dans le contexte de l'étude de fiabilité de circuits électriques, nous pouvons nous appuyer sur un certain nombre de mesures électriques afin de constater s'il y a ou s'il n'y a pas défaillance.

Toutefois, les circuits intégrés sont des circuits électroniques traditionnels construits en un bloc unique. Ainsi une première difficulté s'offre à nous : il apparaît impossible de déterminer la panne du semi-conducteur sans inverser le processus de construction du circuit intégré afin de repérer physiquement le défaut. En détruisant couche après couche les matériaux conducteurs, isolants et semi-conducteurs qui nous interdisent d'observer la structure défaillante, il est possible de voir physiquement la panne : on dit alors qu'on étudie la *physique de la défaillance*. Cependant, une telle opération est longue et coûteuse. Et si par hasard on avait observé plusieurs défauts, il faudrait choisir celui que l'on souhaite observer car cette entreprise de révélation des défauts est exclusivement destructrice.

En pratique, on réalise durant la production une série de mesures métrologiques afin de s'assurer que le circuit est convenablement monté. De plus, on construit à

côté du circuit des structures tests qui sont les témoins de la façon dont une série d'étapes d'intégration s'est déroulée. Ce type de structure témoin permet de vérifier par l'intermédiaire de divers paramètres électriques, le bon fonctionnement d'une partie du circuit intégré. Ces mesures peuvent être répétées un grand nombre de fois (contrairement à l'analyse physique). Ainsi ces structures tests sont utilisées dans l'analyse de la fiabilité du circuit intégré. Par la suite, on conserve un certain nombre de ces structures tests, dites structures critiques, afin de faire des vérifications durant les différentes procédures d'intégration des circuits intégrés.

Pour un circuit intégré, on se doit de tester deux catégories de composants :

1. les éléments actifs (ex. : un transistor, un condensateur, ...) ;
2. les éléments passifs (ex. : une jonction entre deux niveaux, une ligne de métal, ...).

Selon la nature du composant que l'on souhaite tester, on se réfère à différents paramètres électriques afin d'évaluer leur état.

Éléments passifs

Commençons par les éléments passifs qui sont en fait les interconnexions. Leur fonction est simple : elle garantit le passage du courant avec une faible (et constante) résistance. Ainsi, lors du test, nous surveillerons la résistance de la connexion testée au cours de l'expérience. Et lorsque la résistance dépasse un certain seuil, on déclare la panne : voir le graphique 1.6.a représentant l'évolution de résistance au cours du temps. L'augmentation « lente » de la résistance relève d'un comportement typique d'usure de la connexion.

Même si la connexion n'est pas rompue, c'est à dire que la résistance ne tend pas vers l'infini, une variation de résistance engendre de nombreuses répercussions sur les éléments actifs. Si par exemple, notre connexion en état dégradé alimente la grille d'un transistor, alors pour le cas où le transistor est en mode « off », la fonctionnalité du transistor restera la même. Mais pour le cas où le transistor est en mode « on », le passage du courant a toutes les chances de ne pas être autorisé car le courant alimentant la grille arrive avec une intensité moins élevée que prévu. Il y a alors défaillance de la structure, bien qu'elle ne soit pas physiquement détruite. En pratique, on déclare la défaillance d'une connexion lorsque l'augmentation de la résistance varie de plus de 5 à 20 % selon le type de structure testé, la dimension des connexions et les conditions environnementales du test.

Éléments actifs

Maintenant, regardons le cas de tests de défaillance concernant les éléments actifs du circuit, comme par exemple des transistors, des capacités, ou autres ... Dans ce cas, il s'agit de mesurer les propriétés électriques des matériaux qui caractérisent le fonctionnement du composant. Ainsi pour le cas d'un transistor, nous devons mesurer

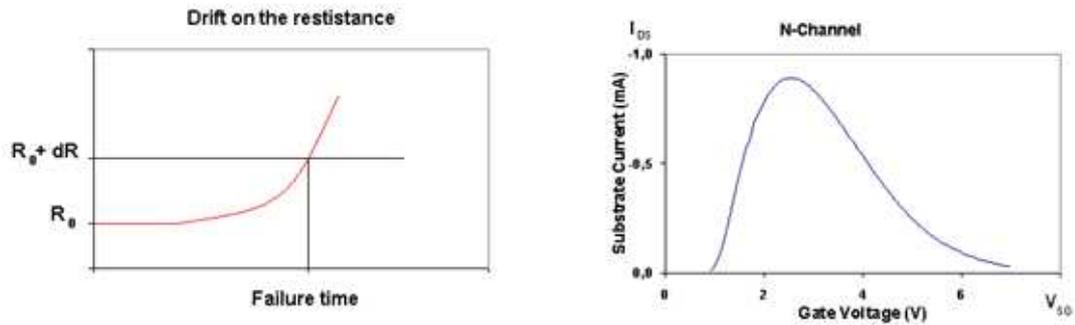


Figure 1.6: En a) évolution de la résistance ; et en b) évaluation des performances d'un transistor durant une période de test.

la tension à appliquer sur la grille afin d'engendrer un champ magnétique qui permettra le passage du courant entre la source et le drain en modifiant les propriétés électriques du semi-conducteur utilisé.

Le graphique de la figure 1.6.b montre l'évolution de l'intensité entre la source et le drain, noté I_{SD} ; lorsque l'on augmente la tension appliquée sur la source et sur la grille, noté V_{SG} . On obtient une courbe concave. Soit \hat{I}_{SD} l'intensité maximale observée, cette valeur caractérise le transistor. Notons \hat{V}_{SG} la tension associée à \hat{I}_{SD} , cette tension est la tension minimale permettant le passage du courant. Si on observe au cours du temps une variation significative de l'intensité \hat{I}_{SD} , on déclare alors que la structure est défaillante comme les règles régissant le passage du courant dans le transistor ont significativement changé.

1.2.2 Modes de défaillance

Maintenant regardons en détails quels sont les principaux modes de défaillance des circuits intégrés. On peut distinguer deux modes de défaillance :

1. Défaillance d'usure continue ;
2. Défaillance d'usure par chocs.

Lors d'une étude de défaillance, il convient de définir précisément le mécanisme physique, chimique ou mécanique de la panne, afin d'utiliser la loi de dégradation la plus adéquate. Ainsi dans le domaine des semi-conducteurs, les chemins de défaillances sont aussi complexes et précis que les procédés utilisés lors de l'intégration du circuit. Par ailleurs, afin de réaliser une image réaliste de la fiabilité des puces microélectroniques, les tests de défaillances sont réalisés sur des structures strictement localisées. Suite à l'analyse approfondie des causes de la défaillance, on peut alors exhiber un mécanisme de défaillance ainsi qu'une (ou plusieurs) variable(s) dite(s) de *stress* liée(s) à la défaillance étudiée.

Le tableau 1.3 dresse une liste non exhaustive de dégradations d'un circuit intégré. Pour chaque type de panne (le type 1 pour usure et 2 pour chocs), on indique la localisation de la panne, le mécanisme de défaillance mis en cause et la (ou les) variable(s) de stress.

Défaut	Type	Détails	Procédé	Variable de stress
Corrosion	1	Contamination de contacts	Dépôt métal-2	Humidité
Corrosion	1	Destruction du di-électrique	Dopage phosphate	Humidité
Corrosion	1	Trou du diélectrique	Dépôt chimique par vaporisation	Humidité
Corrosion	1	Trou du diélectrique	Polissage mécanique et chimique	Humidité
Érosion	1	Micro trou	Photolithographie	Température
Érosion	1	Micro trou	Gravure	Température
Claquage	2	Sputtering	Court-circuit	Tension
Claquage	1	Dépôt d'oxyde par	Trou dans la grille vaporisation	Tension
Radiation	2	Soft error	Passivation par plasma	Rayons Alpha
Migration	1/2	Déformation du sub-strat	Dépôt plasma de l'oxyde	Température
Migration	1	Stressmigration des connexions	Dépôt plasma	Température
Migration	1	Electromigration	Design et dépôt de la connexion	Température et/ou Intensité

Table 1.3: Tableau dressant une liste non exhaustive de défauts en décrivant en détail la cause et le niveau d'intégration.

Ce tableau suggère que le choix du procédé de fabrication est lié de manière très rigoureuse à son emplacement dans la ligne de production, voir fig.1.4. À titre d'exemple, il paraît tout à fait naturel de réserver les procédés sous haute température (à plus de 400 degrés Celsius) pour le FEOL. De même, au cours du BEOL tous les procédés de fabrication sont réalisés à plus faible température : comme la température favorise la formation de zones de déplétion qui amorcent la migration de matériaux dans les lignes conductrices. De plus, le temps des procédés est également réduit, de sorte que les tranches de Silicium sont soumises à moins de stress mécaniques liés à l'augmentation de la température. De même, toute une série de mesures sont recommandées afin d'organiser les procédés physiques, mécaniques et chimiques. Pour plus de précisions, on peut se référer au livre de Wolf (1990 [36]) qui traite en détails de l'intégration de circuits intégrés.

1.3 Défaillance d'usure

Maintenant, regardons un peu plus en détails, les mécanismes de dégradation des semi-conducteurs. Commençons par les mécanismes d'usure. Ensuite, nous regarderons en détails les mécanismes de dégradation liés à l'usure.

1.3.1 Tests d'usure

Dans le domaine de la microélectronique, l'usure est un mode de défaillance naturel induit par la migration de matériaux. Ce phénomène usuel de dégradation touche les surfaces des matériaux semi-conducteurs. En l'occurrence, on le rencontre majoritairement dans les lignes conductrices, et parfois dans les couches isolantes. Dans le reste de ce paragraphe, nous nous limiterons aux migrations métalliques pour deux raisons. D'une part, la migration d'interconnexions est un mode de défaillance critique. D'autre part, les lois de migration de matériaux sont génériques ; ainsi du point de vue du fiabiliste, les modèles associés aux problèmes de migration appartiennent à une même famille de modèles ; à savoir une même distribution de la durée de fonctionnement et les équations de dégradation sont de même forme.

Pour ce qui concerne la migration métallique, les physiciens regroupent ce type de défaillance sous le terme de *metal creeping*. La fatigue des connexions métalliques est engendrée par une lente déplétion des atomes depuis une zone spécifique (définie par des contraintes physiques). À terme, cette migration de matière induit la création d'un trou dans la connexion. Dès lors, le passage du courant dans la connexion est soit significativement réduit, soit interrompu. Dans les deux cas, on déclare la panne.

La densité de courant et la température sont deux variables, dont l'action, éventuellement conjuguée, engendre l'usure de la connexion. De nombreuses publications attestent du rôle de la température dans l'usure, et on nomme ce phénomène le *stressmigration* ; voir McPherson and al. ([24], 1987) . On parle d'*electromigration* ; voir Black ([9], 1969), lorsque le *creeping* est initié sous l'effet du passage du courant ou sous l'effet conjugué du passage du courant et de la température.

Comme nous l'avons vu dans la section précédente, on réalise le constat de la défaillance par le biais de mesures successives de la résistance de la connexion au cours du temps. Dans la pratique, on observe deux configurations dans l'évolution de la résistance, résumées par la figure 1.7.

Tout d'abord, comme sur la figure 1.7.a, la résistance peut croître lentement et de manière continue. Il est utile de remarquer ici, que malgré la création d'un trou dans une ligne de métal le courant passe toujours. En effet, les connexions métalliques sont gainées par une fine ligne de métal qui assure à la fois : la conduction du courant, la protection de la connexion par des contaminants ou des contraintes thermiques et l'adhérence entre les couches conductrices et isolantes. Ainsi la création d'un trou n'induit pas forcément la rupture de la connexion eu égard à ce gainage métallique habituellement nommé *liner*. Donc, selon les propriétés du composant testé, il faut se donner un seuil de résistance à partir duquel on déclare la défaillance. Dans la plupart des cas, on se réfère à un certain pourcentage de croissance de la résistance

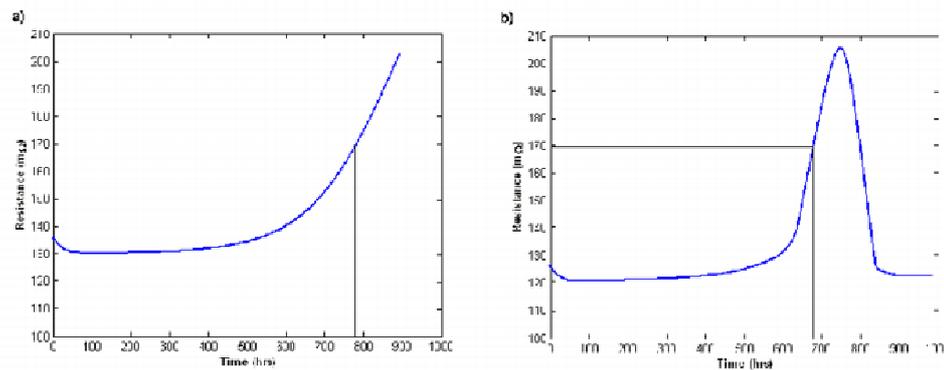


Figure 1.7: Deux configurations dans l'évolution de la résistance durant un test: a) un lent accroissement ; b) un pic de résistance.

entre le temps 0 et le temps t d'observation. En pratique, une croissance de l'ordre de 5 à 20 % de la résistance atteste, de façon significative, de la création d'un défaut de vieillissement.

L'autre cas typique d'évolution de la résistance est représenté sur la figure 1.7.b. Dans ce cas, la résistance augmente très lentement, puis on observe brièvement un pic dans les mesures avant que les mesures ne se stabilisent. Ce cas de figure est rencontré lorsque l'on réalise un test *in situ*, c'est à dire, lorsque le courant est mesuré en continu et qu'il est soit à un niveau usuel, soit à un niveau de stress. Lorsque le trou se crée, la résistance augmente. Puis lorsque le trou se forme et que le courant appliqué à la connexion est relativement élevé, alors la résistance augmente significativement. Et en raison de la petite taille des connexions, le passage du courant produit un arc électrique qui détruit l'intégrité de la structure. Dès lors, cette micro explosion expulse la matière autour du trou et ce conglomerat de matériaux conducteurs et d'oxydes atteint une résistance clampée. Et puis, le pic de résistance fait place à un plateau, comme sur la figure 1.7.b. C'est ainsi que pour ce type de test, on regarde également l'augmentation de la résistance en terme de pourcentages de variation et on définit un seuil de variation de résistance.

Maintenant, développons séparément les principales défaillances d'usure que sont l'*électromigration* et le *stressmigration*.

1.3.2 Électromigration

L'*électromigration* est une défaillance liée à la création d'un trou dans une connexion lors du fonctionnement du composant. La défaillance est donc liée au passage du courant. Dans ce cas, la résistance de la structure sujette à ce défaut augmente jusqu'à la création du trou (cf. fig.1.7.b). Lors du test, on contrôle la résistance de la structure qui nous permet d'observer la panne lorsqu'elle se produit. Cependant, il faut prendre garde car lorsque le trou est finalisé, le passage du courant induit la formation d'un arc électrique au niveau du défaut. Et donc la zone de

panne se retrouve alors disloquée ; dans la plupart des cas elle se ressoude, tout en gardant son caractère non fonctionnelle. Ainsi, les testeurs doivent être capable de repérer ces phénomènes de montée de la résistance qui sont les prémices de ces défauts d'électromigration. Il est donc nécessaire de réaliser des mesures *in situ*. De plus, il est nécessaire d'avoir une grande qualité de mesures, car au-delà d'une certaine marge d'erreur, les variations de résistance sont masquées.

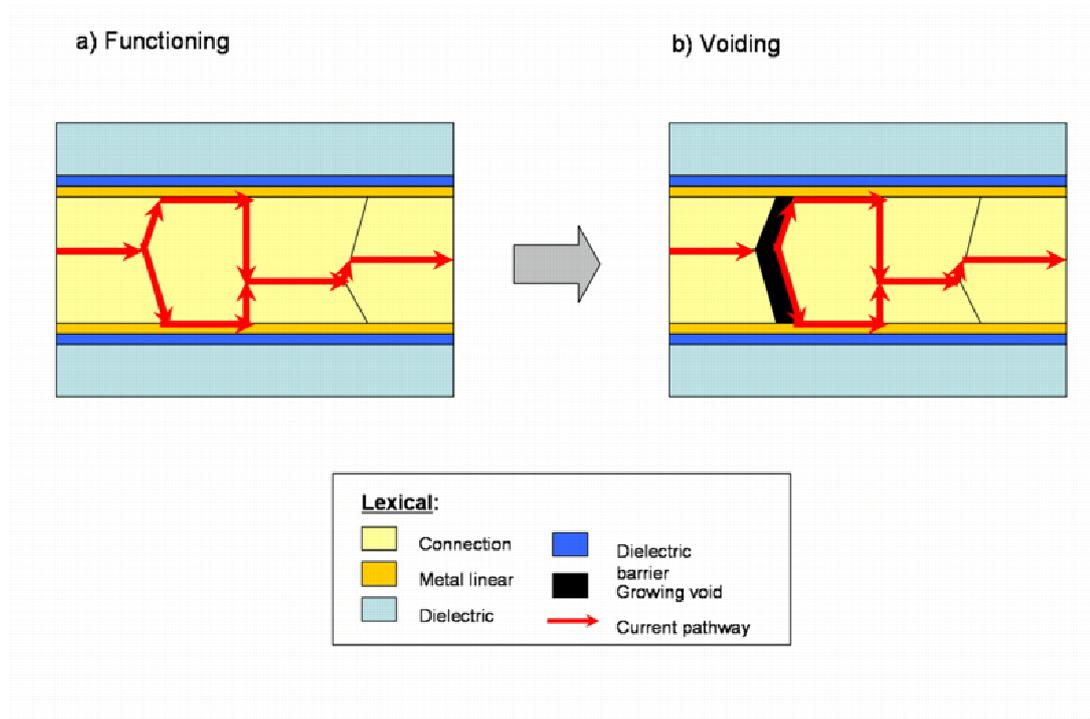


Figure 1.8: Schéma de l'électromigration d'une interconnexion.

La figure 1.8 ci-dessus, illustre le mécanisme de l'électromigration d'une interconnexion. Les flèches rouges représentent le passage du courant, qui se propage entre les particules de métal de la connexion et sur le bord de la ligne de métal. Les flèches noires représentent le flux électronique induit par le passage du courant. Ce flux est d'autant plus élevé que l'intensité du courant est élevée. Cette force provoque la migration des matériaux dans le sens du passage du courant. Ainsi, des zones de déplétion et d'accumulation se créent, et le mécanisme de l'électromigration est alors mis en place jusqu'à la défaillance.

Remarque 1.3.1. Remarquons qu'il y a un effet joule induit par le passage du courant. Ainsi, lorsque la température augmente, l'effet joule est rapidement saturé. Il en résulte un vent électronique d'autant plus fort. La hausse de la température est un paramètre aggravant le phénomène d'électromigration.

1.3.3 Stressmigration

Le *stressmigration* est une défaillance liée à la création d'un trou dans une connexion par des contraintes mécaniques. La migration opère depuis un défaut originel localisé là où les contraintes mécaniques sont les plus fortes, ou bien à un endroit où les contraintes du procédé de fabrication ont créé un défaut dans l'intégrité de la structure. Lorsque la température augmente, les différents matériaux du circuit intégré se dilatent et des contraintes mécaniques se créent dans tout le circuit. En particulier, la région de déplétion est la région dans laquelle les contraintes mécaniques sont plus fortes, et c'est à partir de cette zone que le processus de migration s'engage. Au cours du temps, l'accumulation de la matière produit une lente augmentation de la résistance de la connexion (cf. fig.1.7.a), et ce jusqu'à ce que se crée un trou dans le circuit. En pratique, on n'attend pas que le trou bloque intégralement le passage du courant. On se limite à observer une certaine augmentation de la résistance, afin de repérer la présence d'une large zone de déplétion. Remarquons de plus que la perte de tension liée à l'augmentation de la résistance de certaines connexions peut avoir de grandes conséquences sur le fonctionnement d'un circuit. En effet, gardons en tête que le fonctionnement du transistor dépend de la tension appliquée à ses bornes. Et donc une décroissance de la résistance implique (au mieux) un fonctionnement dégradé.

Le schéma de la figure 1.9 illustre le mécanisme de défaillance du stressmigration d'une interconnexion. La flèche noire représente les forces de tensions mécaniques

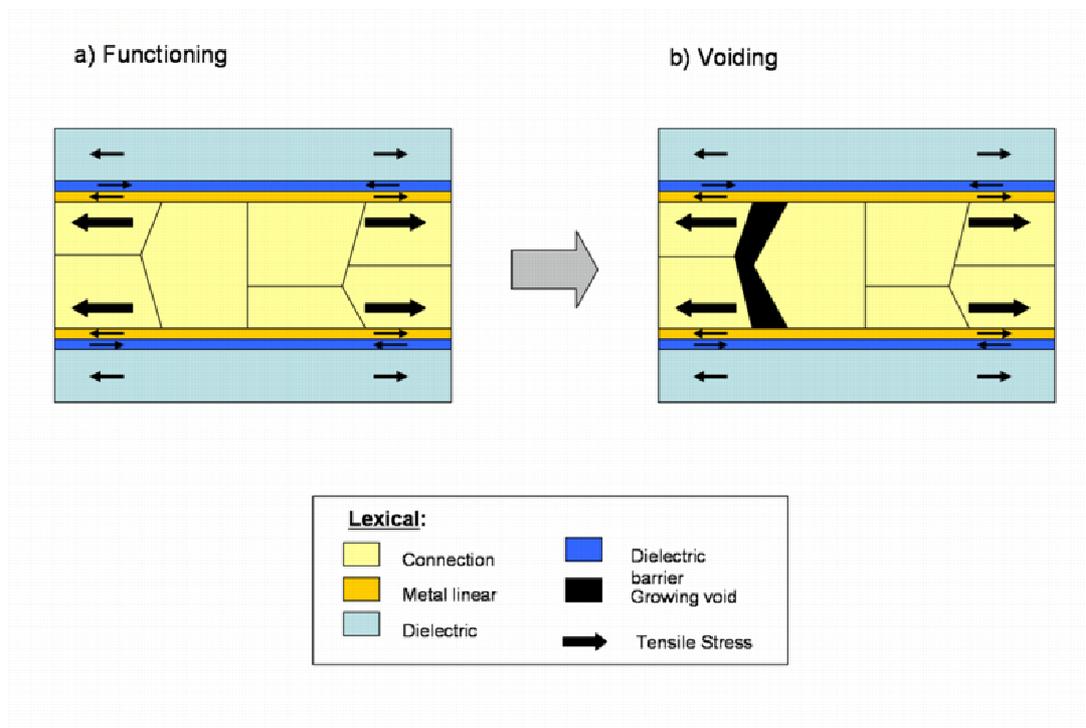


Figure 1.9: Schéma de stressmigration d'une interconnexion.

agissant sur la ligne de métal. Ces forces proviennent de la dilatation des couches de matériaux qui se situent autour de la connexion ; en effet les connexions sont entourées de couches d'oxydes, de barrières d'oxyde ou d'alliages qui ont des caractéristiques thermodynamiques distinctes. Et lorsque les conditions environnementales créent des tensions mécaniques, un phénomène de migration peut apparaître sous des formes comparables à celles rencontrées dans l'électromigration. À savoir que des zones de déplétion et d'accumulation se créent. Ensuite, le mécanisme de migration se propage à partir de ce défaut originel dans le sens des tenseurs mécaniques jusqu'à la défaillance.

1.4 Défaillances par chocs

1.4.1 Test de Défaillance

La cause principale observée en microélectronique des défaillances par chocs est le *claquage diélectrique* (Wolf, 1990 [36]). Il s'agit de la panne naturelle des composants actifs comme les transistors, les condensateurs, etc ... En résumé, le claquage diélectrique provient d'une lente accumulation d'électrons ou d'ions dans une région isolante ou semi-conductrice. Leur présence est accélérée par un changement des conditions thermodynamiques. Elle induit un champ électromagnétique qui modifie les propriétés électroniques du composant. À terme, une zone isolante devient semi-conductrice et le passage du courant produit un court-circuit.

Remarque 1.4.1. Remarquons que ce type de panne peut également toucher les couches isolantes protégeant les deux connexions contiguës. Cependant, en modifiant l'architecture des circuits intégrés modernes, ce type de panne est devenu marginal.

Ces défaillances sont caractérisées par la mesure des paramètres électriques des composants. Dans le cas des transistors, on réalise deux mesures :

1. Tension entre la grille et la source du transistor ;
2. Intensité entre la source et le drain.

Soient V_{GS} la tension mesurée entre la grille et la source du transistor et I_{DS} l'intensité mesurée entre la source et le drain. Rappelons brièvement le fonctionnement du transistor : le courant passe entre la source et le drain lorsque la grille est suffisamment alimentée. Donc, lorsque l'on augmente la tension appliquée à la grille, la densité de courant mesurée entre la source et le drain augmente jusqu'à un seuil optimal ; notons \hat{I}_{DS} l'intensité maximale du transistor. Ce paramètre caractérise les propriétés électriques du transistor dérivées des propriétés électriques du canal (constitué de matière semi-conductrice dopée).

Ce phénomène est représenté sur la figure 1.10, sur laquelle on peut voir : la tension V_{GS} en fonction de I_{DS} pour un transistor fonctionnel et pour un transistor défaillant. \hat{I}_{DS} , qui caractérise le transistor, est l'intensité maximale du transistor. On note que la forme des deux courbes est comparable, cependant leurs \hat{I}_{DS} et leurs échelles sont différentes. En pratique, pour mesurer l'évolution des propriétés du transistor, on

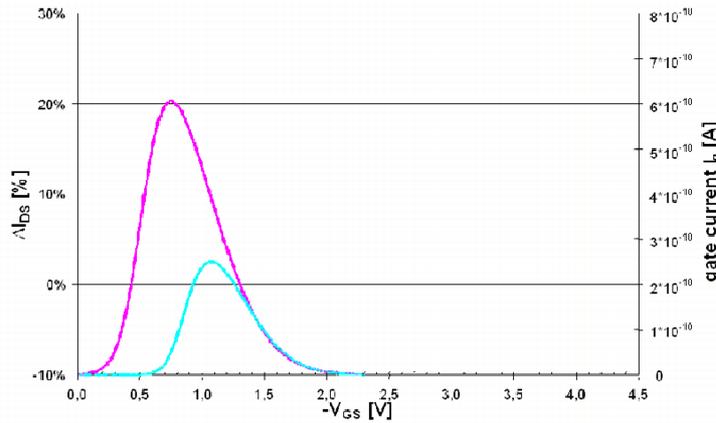


Figure 1.10: Relation entre la tension source-grille et la variation de l'intensité drain/source ; la courbe violette représente le transistor en fonctionnement usuel et la courbe bleu clair montre le transistor en fonctionnement dégradé.

mesure \hat{I}_{DS} l'intensité maximale du transistor et on regarde la différence d'intensité maximal ΔI_{DS} entre le temps de mesure t et la valeur initiale. Ainsi, lorsque cette différence est significative (pour un seuil de signification donné par l'expérience), on peut repérer le changement des propriétés du transistor et par là même, on constate le dysfonctionnement du transistor.

Remarque 1.4.2. D'une manière générale, on teste la fiabilité d'un composant actif en étudiant les variations de ses propriétés électriques au cours du temps. Ainsi, lorsqu'on étudie un transistor, on suit l'évolution de \hat{I}_{DS} l'intensité drain-source maximale ; lorsqu'on étudie la fiabilité d'un condensateur, on suit l'évolution de la capacité au cours du temps ... etc.

1.4.2 Claquage d'un transistor

Maintenant que nous avons vu comment repérer le *claquage*, regardons plus en détails ce mécanisme de défaillance. Pendant son fonctionnement, le passage du courant peut provoquer d'une part l'accumulation d'électrons dans la grille du transistor, et d'autre part, une accumulation d'ions dans le canal (la zone semi-conductrice située entre la source et le drain voir fig.1.1). Ce phénomène dual est à l'origine des défaillances observées dans les transistors. L'accumulation de particules chargées provoque un changement dans les caractéristiques du transistor. Et puis à terme, un court-circuit se produit au niveau de la grille, et ce phénomène détruit en général les matériaux présents entre la grille et le drain du transistor.

Le schéma de la figure 1.11 illustre le claquage diélectrique. La flèche rouge essais standards représente le passage du courant. Lors de chaque mise en fonctionnement (à chaque choc), le passage du courant peut provoquer l'accumulation d'électrons dans la grille du transistor ; de fait le flux du champ magnétique qui agit sur la zone semi-conductrice se trouve modifié par cette accumulation. Ainsi on peut apprécier l'évo-

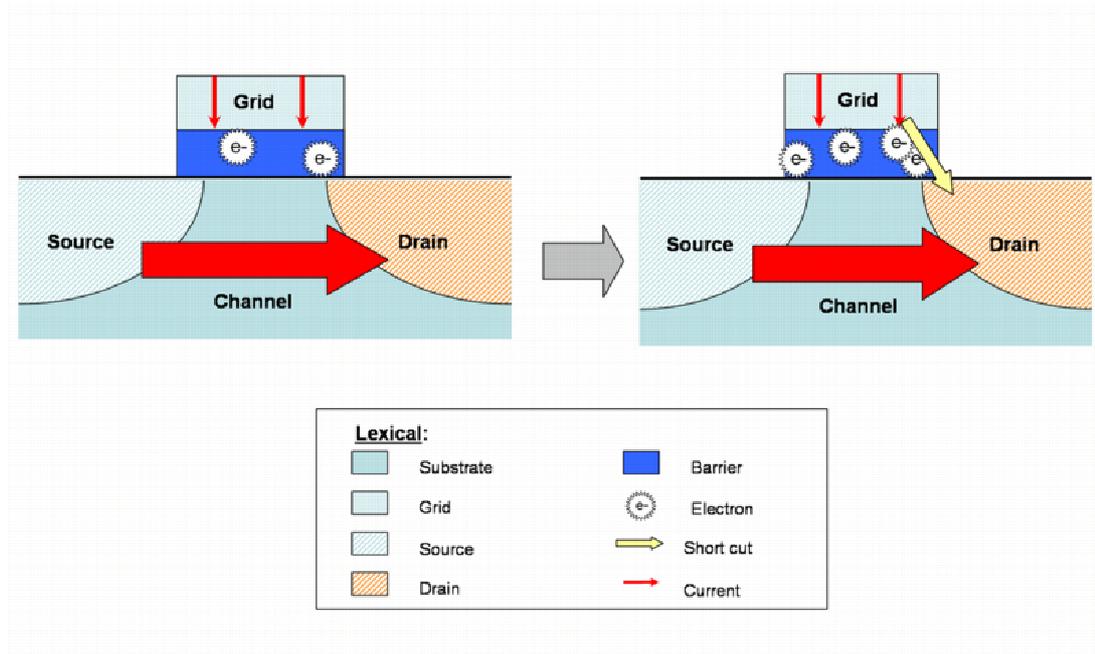


Figure 1.11: Schéma du claquage diélectrique d'un transistor.

lution de la défaillance en mesurant la différence ΔI_{DS} de l'intensité maximale du transistor. Et puis à terme, comme sur le schéma (1.11.b), l'accumulation provoque un court-circuit entre la grille et le drain.

1.5 Généralités sur la fiabilité en microélectronique

1.5.1 Retour d'expérience

La performance de fiabilité de tout composant industriel doit être estimée régulièrement, afin de s'assurer que la fiabilité est conforme aux normes des organismes d'homologation, d'améliorer le rendement par une politique de maintenance préventive et de mettre en place une politique de qualité auprès des clients.

Le retour d'expérience (REX) est une méthodologie construite afin de répondre à ces enjeux de qualité. Elle a pour objectif de collecter, d'archiver, d'analyser les données de défaillance et de dégradation. Lorsque l'on réalise la collection de données statistiques d'un composant mature, il survient en général différentes contraintes : d'une part la censure des données, liée au fait que toutes les unités tests ne tombent pas en panne durant l'expérience ; et d'autre part, les données peuvent avoir un (ou plusieurs) comportement(s) masqué(s), ce qui se produit par exemple lorsqu'il existe plusieurs modes de défaillances.

Le REX permet d'utiliser les données antérieurement collectées afin de définir des lois à *priori* des paramètres du modèle de fiabilité. En utilisant l'inférence de Bayes et

les données issues du REX, un grand nombre de travaux ont montré comment estimer les caractéristiques de fiabilité à partir de données censurées ou manquantes (Bacha et al. [2]).

1.5.2 Standards de fiabilité

Depuis plus d'un demi siècle, le monde industriel a mis en commun un certain nombre d'informations issues du retour d'expérience dans le but de réaliser des standards. Par l'intermédiaire de cette politique de qualité, les procédés industriels disposent d'un grand nombre d'informations à *priori*. Le standard le plus utilisé est le U.S. Military Handbook, mais il en existe bien d'autres comme le Bellcore RPP, le British Telecom RHD ou encore les procédures NTT, CNET, Siemens ou LM Ericsson's, le ATT Handbook ...

Dans ces standards, on retrouve une liste exhaustive des caractéristiques de fiabilité liés à toute une gamme de composants microélectroniques dépendant du degré d'intégration (taille du support, trait, ...), des dimensions des structures, des propriétés des matériaux, de la fonction du composant (ex. logique ou mémoire, ...), une série de structures critiques (contacts en série, éléments logiques, ...). Elles proposent également une série de tests de référence qui permettent de garantir la qualité des caractéristiques de fiabilité en fonction de différents mécanismes de défaillances. Cependant l'utilisation de ces standards est limitée aux technologies matures. À chaque instant, ces standards doivent évoluer en fonction des changements technologiques. C'est pourquoi on a alors recours à des essais de fiabilité.

1.5.3 Essais de fiabilité et accélération

Les essais standards sont dans une impasse lorsque l'on étudie la fiabilité d'un composant à longue durée de vie sans information à *priori*. C'est pourquoi les fiabilistes mettent en place des essais accélérés. En électronique, on cite souvent un accident de stockage lorsqu'on souhaite illustrer l'importance des tests accélérés. À la suite d'une exposition prolongée sous température caniculaire, un stock complet de condensateurs est tombé en panne. De ce type d'incident, on peut tirer deux enseignements : en premier lieu, on a pris conscience de l'importance des conditions de stockage des composants électroniques et du conditionnement du produit. Et en second lieu, il montre qu'en modifiant les paramètres environnementaux, il était possible d'accélérer le processus de défaillance d'un composant électronique.

Ainsi, lors d'un essai accéléré, on étudie la dégradation d'une série d'unités de tests placées sous différents environnements de stress en vue d'accélérer l'apparition des défauts. Les essais accélérés sont plus complexes à mettre en place et à exploiter que les essais standards. Mais ils ont un double avantage, ils permettent de :

1. Optimiser la durée des essais de fiabilité ;
2. Améliorer la connaissance du phénomène de dégradation.

1.6 Modèles de fiabilité

1.6.1 Modèle et notations

Maintenant discutons des modèles de fiabilité. Soit Z une variable aléatoire représentant la *durée de vie* d'un produit. Le but de l'étude de fiabilité est d'estimer la *fonction de fiabilité*, notée $R(\cdot)$. Elle est définie comme suit:

$$R(t) = P(Z > t)$$

où la fonction $R(\cdot)$ est définie pour tout $t \in \mathbb{R}^+$. Supposons de plus que cette loi de Z est paramétrique, notée $\mathcal{L}(\theta)$, et posons θ le paramètre de cette loi.

Dans le domaine des semi-conducteurs (Nelson [27] ou Meeker et Escobar [25]), on considère que la distribution des durées de vie suit une loi du type *log-location scale*. De tels modèles paramétriques s'écrivent de manière très simple. Fixons $Y = \log(Z)$, on considère alors le problème de régression suivant:

$$Y = \mu + \sigma\varepsilon,$$

où les éléments définissant le modèle sont :

- μ : le paramètre de *localisation* ;
- σ : le paramètre *d'échelle* ;
- ε : une variable aléatoire représentant les *erreurs*.

Le choix de la distribution de ε vient directement du mécanisme de dégradation du composant étudié.

Dans la section précédente, nous avons cité les deux types de dégradation majoritaires opérant sur les semi-conducteurs :

- Usure continue ;
- Usure par chocs.

Dans les sections suivantes, nous discuterons des lois de probabilité classiques utilisées en fiabilité : la loi Lognormale et la loi de Weibull. Nous regarderons plus en détails, les raisons pour lesquelles les fiabilistes utilisent généralement ces deux lois de probabilité.

1.6.2 Distribution Lognormale

Une variable aléatoire Z suit une loi Lognormale de paramètres $\theta = (\mu, \sigma)$, lorsque $Y = \log(Z)$ et Y est une variable aléatoire Gaussienne de moyenne μ et de variance σ^2 . Ainsi, la loi Lognormale est une simple transformation logarithmique d'une loi Normale.

C'est ainsi que toute une série de propriétés asymptotiques de la loi Normale et de ses estimateurs reste valable dans le cas Lognormale. Cette loi est couramment utilisée dans les semi-conducteurs, mais aussi dans les sciences économiques ou les études médicales. Toutefois, la loi Lognormale n'est pas qu'une loi issue d'une transformation de variable en vue d'une normalisation. En fiabilité cette loi a un sens souvent associé à la durée de vie de composants mécaniques.

Soit $Y = \log(Z)$ définie telle que la loi de Y est Normale de moyenne μ et de variance σ^2 ; alors ϕ la fonction de densité de Y est telle que:

$$\phi(t) = \frac{1}{t\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(\log(t) - \log(\mu))^2}{2\sigma^2}\right), \quad \forall t > 0.$$

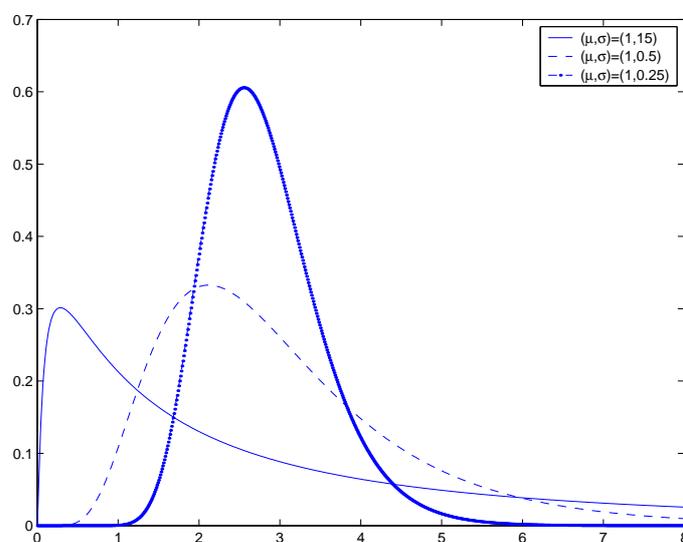


Figure 1.12: Représentation de la fonction de densité de probabilité ϕ d'une loi Lognormale pour différentes valeurs du paramètre de forme $\sigma = 1,5 ; 0,5 ; 0,25$ et $\mu=2$.

Notons que la fonction de densité de probabilité ϕ offre une grande variété de formes. Cette variété est visible par exemple sur la figure (1.12) qui représente ϕ pour différentes valeurs du paramètre de forme σ .

La loi Lognormale est la loi de durée de vie naturelle associée aux mécanismes de dégradation mécanique. Et nous allons présenter, au travers d'un exemple typique (voir le livre de Bon, [10]), quelles hypothèses concernant le mécanisme de dégradation mènent à supposer que la loi de la durée de vie d'un composant est Lognormale. Supposons que nous étudions la durée de vie d'une poutre sujette à des sollicitations mécaniques. La répétition des effets mécaniques affecte un défaut de structure de la poutre, jusqu'à la rupture de la structure. Notons X_0 la largeur du défaut initial. Suite à chaque sollicitation mécanique nous mesurons la largeur du défaut, et nous notons X_1, X_2, \dots, X_n les variables aléatoires des n mesures successives de ces largeurs. À terme, la structure fragilisée par l'accumulation d'un grand nombre n de sollicitations cède et on observe la défaillance.

À la suite de i sollicitations, la largeur du défaut augmente de $(X_i - X_{i-1})$. Supposons de plus que la largeur de l'impact de i sollicitations est proportionnel à X_i , alors notons C_i la variable aléatoire, liée à l'évolution de la fissure, telle que $(X_i - X_{i-1}) = C_i \cdot X_i$. Ainsi au bout de n sollicitations, on constate que :

$$\sum_{i=1}^n C_i = \sum_{i=1}^n \frac{(X_i - X_{i-1})}{X_i}.$$

Et en supposant que l'évolution de la fissure est suffisamment lente, on peut donc utiliser l'approximation suivante:

$$\begin{aligned} \sum_{i=1}^n C_i &= \int_{X_0}^{X_n} \frac{dX}{X} \\ &= \log(X_n) - \log(X_0). \end{aligned}$$

Ainsi, on obtient une approximation du $\log(X_n)$, comme la somme de variables aléatoires $(C_i)_{i=1, \dots, n}$:

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

Par la loi des grands nombres, le logarithme de la largeur de X_n est asymptotiquement une loi normale. C'est ainsi qu'on associe naturellement la loi de distribution Lognormale aux défaillances d'usure mécanique.

Pour le cas des semi-conducteurs, les phénomènes d'usure sont observés dans les connexions. Dans les articles de référence de la physique de la défaillances des circuits intégrés Black [9], McPherson et al. [24] ou plus récemment Ogawa et al. [29], on constate que la dégradation des connexions est due à la migration de masse induite à la suite d'efforts mécaniques. Suite à la migration de masse, une zone de déplétion et une zone d'accumulation se créent à la faveur d'une jonction reliant deux niveaux métalliques (au dessus, au dessous ou dans la jonction). Lorsque la résistivité de cette zone de déplétion croît à mesure que le trou grossit, ce phénomène est connu sous le nom de stressmigration (voir 1.3.3). De plus ce phénomène obéit à des règles mécaniques du même ordre que pour notre exemple de fiabilité de poutres. Notons que l'électromigration (voir 1.3.2) relève d'un mécanisme de défaillance comparable ; mais dans ce cas précis, la migration est induite par le passage du courant plutôt que par des tenseurs mécaniques. Ainsi la loi de durée de vie d'un composant sujet au stressmigration est la loi Lognormale.

1.6.3 Distribution Weibull

La fonction de répartition de Weibull est utilisée dans un grand nombre de cas où le mécanisme de dégradation est complexe. Le domaine d'application de la loi de Weibull est très varié : elle va de l'électronique aux phénomènes mécaniques. Un grand nombre d'ingénieurs utilisent cette loi en raison de la forme de son taux de dégradation en puissance qui s'ajuste dans un grand nombre de cas pratiques. Plus généralement,

lorsque le taux de dégradation est une fonction *complexe*, on utilise naturellement cette loi de distribution. La loi de Weibull est l'une des trois lois des *valeurs extrêmes* définies par Gumbel (Coles [12], chap.3). Les lois des valeurs extrêmes sont des lois naturelles en fiabilité, et, parmi elles, la loi de Weibull est la plus utilisée.

Comme pour le cas Lognormal, regardons dans un cas pratique les hypothèses d'application de la loi Weibull. Prenons l'exemple de la corrosion d'une surface plane. Soit X_1, \dots, X_N les N variables aléatoires *i.i.d.* représentant la profondeur de N marques de corrosion observées sur une surface plane. Afin d'étudier le phénomène de corrosion, il nous faut repérer le trou le plus profond parmi N trous sur la surface testée. Notons $U = \sup_{j=1, \dots, N} X_j$ la variable aléatoire associée au trou le plus profond. La loi de U dépend de la loi jointe des $(X_j)_{j=1, \dots, N}$ et de N . Lorsque N tend vers l'infini, la loi de U est par définition une loi des valeurs extrêmes.

On dénombre trois lois des valeurs extrêmes :

- Type I : La loi de Gumbel (ou SEV pour Smallest Extreme Value) est la loi des valeurs extrêmes associée au cas où la loi jointe des (X_j) est Gaussienne, exponentielle ou logistique;
- Type II : La loi de Fréchet est la loi des valeurs extrêmes associée au cas où la loi jointe des (X_j) suit une loi de Cauchy ;
- Type III : La loi de Weibull est la loi des valeurs extrêmes associée au cas où la loi jointe des (X_j) est une fonction non dégénérée et à support de variation borné.

Maintenant que les hypothèses menant à la loi de Weibull sont exposées ; regardons les propriétés de cette loi. Soit T une variable aléatoire de loi de Weibull $W(\beta, \eta)$ où les paramètres β et η définissent ψ la fonction de densité de T telle que :

$$\psi(t) = \beta \eta^{-\beta} t^{\beta-1} \exp(-(t/\eta)^\beta).$$

Le paramètre η est le paramètre d'échelle, il caractérise l'échelle de variation de T . Et le paramètre β est le paramètre de forme. La figure 1.13 illustre les différentes formes de ψ . Ce paramètre positif β caractérise la forme de la fonction de densité de sorte que :

- si $\beta < 1$, la fonction de densité décroît de l'infini à zéro;
- si $1 < \beta < 2$, la fonction de densité a une tangente verticale en zéro, elle croît vers un maximum local et puis elle décroît ;
- si $\beta > 2$, la fonction de densité a une tangente horizontale en zéro et deux points d'inflexions.

L'espérance de T vaut :

$$E(T) = \eta \Gamma(1 + 1/\beta),$$

où Γ désigne la fonction gamma d'Euler. Et la variance de T vaut :

$$Var(T) = \eta^2 \left[\Gamma\left(1 + \frac{2}{\beta}\right) - \Gamma^2\left(1 + \frac{1}{\beta}\right) \right].$$

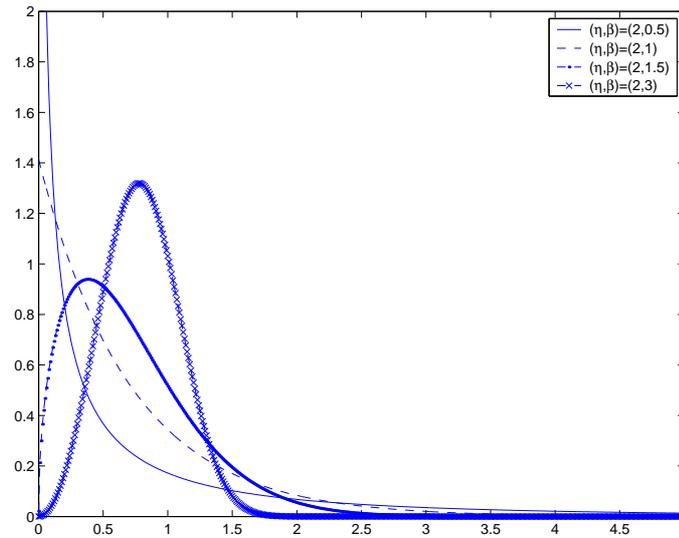


Figure 1.13: Représentation de la fonction de densité de probabilité ψ d'une loi Weibull pour différentes valeurs du paramètre de forme $\beta = 0,5 ; 1 ; 1,5 ; 3$ et $\eta=2$.

Remarque 1.6.1. Notons que la loi de Weibull est connue pour être une loi à densité non exponentielle, sauf lorsque le paramètre β vaut 1 ; auquel cas, la loi de T est la loi exponentielle.

Dans le domaine des semi-conducteurs, l'un des modes de défaillance majoritairement rencontré est la défaillance suite à *l'accumulation de chocs*. Ce type de défaillance est observé sur des structures actives, comme les transistors ou des condensateurs. Et on utilise habituellement la loi de Weibull pour modéliser la fiabilité de ce genre de défaut. Le principe de ce mode de défaillance est le suivant. Suite à l'accumulation de chocs liés du passage du courant, les régions dopées sont sujettes à l'accumulation d'électrons ou de trous d'électrons selon leur charge (P ou N). Ainsi sur la surface dopée, certaines zones sont *contaminées* par des électrons ou des trous d'électrons. On se retrouve alors dans un cas de figure comparable à celui de la corrosion, où N zones d'une surface plane sont attaquées par la corrosion. Et lorsque N tend vers l'infini et en vertu des connaissances concernant les atomes, il paraît raisonnable de supposer que la distribution du nombre d'électrons accumulés pour un choc est une fonction non dégénérée et à support de variation borné (comme il y a un nombre fini et contrôlé d'ions dans le canal), alors on retrouve la loi de Weibull.

1.7 Modèle accéléré

1.7.1 Hypothèses et notations

Soit $Y = \log(Z)$ une variable aléatoire représentant le logarithme des durées de vie, considérons le problème de régression suivant :

$$Y = \mu + \sigma\varepsilon,$$

où les éléments définissant le modèle sont :

- μ : le paramètre de *localisation* ;
- σ : le paramètre *d'échelle* ;
- ε : une variable aléatoire représentant les *erreurs*.

Le modèle accéléré est utilisé afin d'étudier la fiabilité d'un composant test à partir d'essais accélérés, noté ALT (pour Accelerating Life Test). Le principe des essais accélérés est de placer les unités de test sous différents environnements de stress en vue d'accélérer l'apparition des défauts.

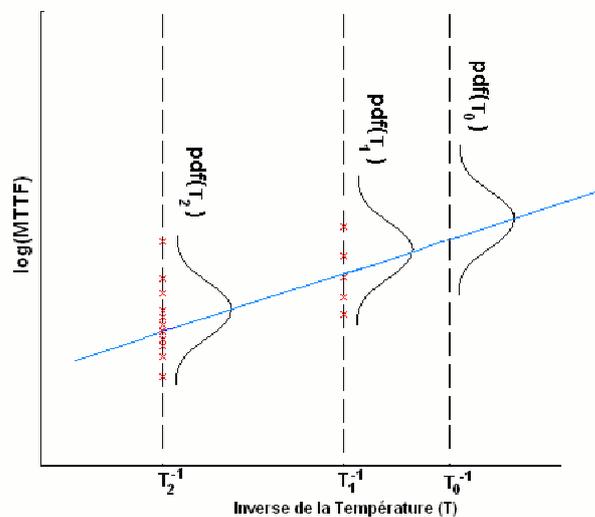


Figure 1.14: Principe du test ALT.

Ce cadre très spécifique d'un modèle accéléré mène à considérer deux hypothèses fortes ([23], chapitre 9). Les hypothèses d'ALT sont les suivantes :

Hypothèse 1.1.

(H1) La sévérité des niveaux d'intensité de stress (caractérisé par les covariables x) ne change pas la distribution des observations Z (ou Y la log-durée de vie), soit $\Phi(t) = P\left(\frac{Z-\mu}{\sigma} \leq t\right)$ la distribution de Z , mais l'environnement de stress a une influence sur les paramètres du modèle, à savoir $\theta = (\mu, \sigma)$.

(H2) La relation de stress f reliant μ et les covariables \mathbf{x} est connue :

$$\mu = f(\mathbf{x}; a_0, a_1, \dots, a_{r-1}),$$

mais les paramètres de stress a_0, a_1, \dots, a_{r-1} ne le sont pas ; de plus la relation d'accélération n'est valable que pour un domaine environnemental donné.

Le but de ce modèle est d'estimer μ en tant que fonction définie par a_0, a_1, \dots, a_{r-1} à partir d'un échantillon (Z, \mathbf{x}) . La première hypothèse (H1) concerne le processus de dégradation. Il suppose que les caractéristiques de fiabilité sont sujets à l'environnement, caractérisé par \mathbf{x} . Par ailleurs, il affirme que les variations environnementales n'influent pas sur le processus de dégradation, et donc la loi de distribution de la durée de vie reste la même quelque soit les conditions de stress \mathbf{x} . Si cette hypothèse (H1) n'était pas faite ; nous ne pourrions réaliser aucune interpolation entre deux jeux de données obtenus sous deux conditions environnementales différentes.

En second lieu, (H2) émet l'hypothèse qu'il existe une relation f entre la durée de vie et les conditions de stress. La fonction d'accélération f décrit la manière dont des différents environnements de stress opèrent sur la distribution des durées de vie, via μ . Cette fonction d'accélération est choisie en fonction du mécanisme physique du vieillissement du composant testé. Ainsi un modèle accéléré est entièrement défini par Φ , la distribution de la durée de vie, et par une fonction d'accélération f reliant la durée de vie aux conditions environnementales. Lorsque f est linéaire, on parle alors de modèle Accelerated Failure Time ou modèle AFT.

Remarque 1.7.1. Pour le cas particulier où la covariable \mathbf{x} est un simple scalaire, on considère un intervalle que la relation est valable sur un certain intervalle de stress $[\mathbf{x}_{\min}; \mathbf{x}_{\max}]$. Remarquons que l'hypothèse concernant $[\mathbf{x}_{\min}; \mathbf{x}_{\max}]$ permet de nous assurer que d'une part pour un certain *niveau d'intensité* du stress du test, la relation est connue. Et d'autre part, elle nous assure que les inférences concernant la fiabilité du composant test sont toujours valables pour l'environnement opérationnel \mathbf{x}_0 . Cette hypothèse tient son origine des propriétés thermodynamiques des matériaux. En effet, les comportements mécaniques des matériaux varient en fonction de leur environnement, comme les états des matériaux changent en fonction de la pression atmosphérique et de la température.

Remarque 1.7.2. Par ailleurs, notons que dans un certain nombre de cas partiques, le paramètre σ dépend également des conditions environnementales. Dans ce cas la relation de (H2) devient:

$$\theta = (\mu, \sigma) = f(\mathbf{x}; a'_0, \dots, a'_{r-1}).$$

Dans le domaine de la microélectronique le paramètre σ peut décroître avec la température dans le cadre de certaines technologies Aluminium (voir travaux de Schwarz, [34]), cependant ce n'est pas le cas le plus général. En effet n'est pas généralisable aux connections en cuivres qui sont maintenant habituellement utilisés (voire les travaux d'Ogawa et al., [29]). C'est pourquoi dans nos trvaux nous avons considéré que le paramètre σ ne dépend pas du stress.

Le modèle d'accélération est illustré par la figure (1.14). Sur l'axe des ordonnées, on a placé les observations (ici le logarithme de la durée de vie) et sur l'axe des abscisses, on a placé les variables environnementales associées à chaque observation ; pour cet exemple il s'agit de l'inverse de la température. La droite bleue représente la relation liant la durée de vie moyenne et l'inverse de la température par une formule de type Arrhénius. Et pour une température \mathbf{x}_0 fixée, on a tracé la fonction de densité de

probabilité du logarithme de la durée de vie pour cet environnement de stress. Ainsi, dès lors que la fonction d'accélération est avérée, on a affaire à un modèle paramétrique standard. Par contre, lorsque la relation d'accélération est inconnue, alors on a affaire à un modèle non paramétrique ([23], p.447, ou [27] chap.10).

1.7.2 Accélération classique

Le modèle d'accélération tend à représenter la complexité du mécanisme de vieillissement par rapport aux conditions environnementales de stress. La physique de la défaillance consiste à étudier physiquement les origines de la panne, et ainsi elle conduit à décrire la cinétique du mécanisme de vieillissement à l'aide d'équations physiques, mécaniques, thermodynamiques ou chimiques. Dans le modèle de fiabilité, nous avons vu que pour la relation d'accélération f lie le paramètre de localisation du modèle aux conditions de stress, comme dans l'hypothèse (H2) :

$$(H2) \quad \mu = f(\mathbf{x}).$$

Le choix de l'accélération f est effectué à partir de la physique de la défaillance. La fonction f est choisie dans le but de retranscrire la cinétique du phénomène physique, mécanique, chimique ou thermodynamique auquel obéit le composant étudié. Ainsi, on déduit de la physique de la défaillance les fonctions d'accélération spécifiques à la plupart des cas pratiques. Elles relient un facteur d'accélération, noté F_A , au stress \mathbf{x} . Notons que dans le cas des données Lognormales, le facteur d'accélération est égal à le MTTF (mean time to failure), de sorte que le $\log(F_A)$ vaut le paramètre de localisation μ . Et dans le cas des données Weibull, le facteur d'accélération est égal au pourcentage d'ordre 63.2%, de sorte que le F_A vaut le paramètre de localisation du modèle.

Nous proposons, ci-dessous, une liste non exhaustive de relations d'accélération classiques [9, 27, 23, 25, 29] :

- *Règle de puissance* : $F_A(x) = C.\mathbf{x}^{-p}$ où \mathbf{x} est une variable environnementale univariée, et (C, p) sont les paramètres d'accélération du modèle ;
- *Arrhénius* : $F_A(\mathbf{x}) = Q.\exp(-E_A/\kappa.\mathbf{x})$ où \mathbf{x} est une variable environnementale univariée, et (Q, E_A) sont les paramètres d'accélération du modèle ;
- *Eyring* : $F_A(\mathbf{x}) = \mathbf{x}.\exp(A - B/\mathbf{x})$ où \mathbf{x} est une variable environnementale univariée, et (A, B, C, D) sont les paramètres d'accélération du modèle ;
- *Eyring généralisé* : $F_A(T, V) = A.T.\exp(-B/\kappa.T + C.V + D.V/T)$ où $\mathbf{x} = (T, V)'$ est une variable environnementale bivariée, et (A, B, C, D) sont les paramètres d'accélération du modèle ;
- *Black* : $F_A(T, j) = j^{-n}.Q.\exp(-E_A/kT)$ où $\mathbf{x} = (T; j)$ est une variable environnementale bivariée, et (n, Q, E_A) sont les paramètres d'accélération du modèle.

PS : ici κ est une constante universelle, il s'agit de la constante de Boltzmann qui vaut $8,61.10^{-5}eV/^\circ K$.

1.7.3 Modèle standard de vie accélérée

La classe des modèles standards de vie accélérée (SVA) est une classe de modèles accélérés qui s'applique à un grand nombre de cas pratiques. Par exemple, le modèle accéléré modèle AFT est un modèle SVA. Les sections précédentes présentent les hypothèses d'accélération selon un angle appliqué, et dans cette section nous proposons une approche plus technique. Par ailleurs, nous avons ainsi l'occasion de parler brièvement du cas où les stress changent durant l'expérience accélérée. Pour rédiger cette présentation du modèle SVA, nous nous sommes référés à l'article de Bagdonavicius, Gerville-Réache, Nikoulina et Nikulin [3] qui propose une analyse statistique complète du modèle SVA.

Dans un souci de généralisation, nous considérerons ici que les stress \mathbf{x} peuvent être multidimensionnels et qu'ils peuvent évoluer au cours du temps. Ainsi :

$$\mathbf{x} = \mathbf{x}(\tau), \tau \geq 0, \quad \text{où } \mathbf{x} : [0; \infty[\rightarrow \mathcal{X} \subseteq \mathbb{R}^{r-1}.$$

Soit Z la durée de vie sous le stress $\mathbf{x}(\cdot)$ est vue comme une variable aléatoire non négative et absolument continue de fonction de fiabilité $R_{\mathbf{x}(\cdot)}(t) = P_{\mathbf{x}(\cdot)}(Z > t)$.

Pour exploiter des données issues d'un essai accéléré, il faut construire un modèle qui décrit la façon dont $R_{\mathbf{x}(\cdot)}$ (ou une autre fonction caractéristique de Z) évolue en fonction du stress.

Soit $\mathcal{X}_0 \subset \mathcal{X}$ un sous-ensemble de stress constants (définissant les conditions opérationnelles), soit $\mathbf{x}_0 \in \mathcal{X}_0$ et soit $R_{\mathbf{x}_0}^{-1}(p) = \inf\{s : R_{\mathbf{x}_0}(s) \geq p\}$ la fonction inverse de $R_{\mathbf{x}_0}$. La fonction $v_{\mathbf{x}(\cdot)}(t) = R_{\mathbf{x}_0}^{-1} \circ R_{\mathbf{x}(\cdot)}(t)$ est la fonction de ressource utilisée jusqu'au temps t . Et donc $\forall \mathbf{x}(\cdot) \in \mathcal{X}$:

$$P_{\mathbf{x}_0}(Z \geq v_{\mathbf{x}(\cdot)}(t)) = P_{\mathbf{x}(\cdot)}(Z \geq t).$$

Ainsi le modèle SVA est défini sur l'ensemble de stress \mathcal{X} si il existe un facteur d'accélération $F_A : \mathcal{X} \rightarrow \mathbb{R}_+$ tel que :

$$\forall \mathbf{x}(\cdot) \in \mathcal{X} : \quad \frac{d}{dt} v_{\mathbf{x}(\cdot)}(t) = F_A(\mathbf{x}(t)). \quad (1.1)$$

Le modèle défini par (1.1) émet l'hypothèse que la vitesse d'utilisation de la ressource à l'instant t ne dépend que de la valeur du stress appliqué à cet instant, ainsi :

$$R_{\mathbf{x}(\cdot)}(t) = R_{\mathbf{x}_0} \left(\int_0^t F_A(\mathbf{x}(\tau)) d\tau \right).$$

De sorte que nous avons un lien propre à définir une relation d'accélération entre $R_{\mathbf{x}(\cdot)}$ et $R_{\mathbf{x}_0}$. Lorsque les conditions de stress changent au cours de l'expérience, cette relation d'accélération reste valide. Et lorsque les conditions de stress ne changent pas $R_{\mathbf{x}}(t) = R_{\mathbf{x}_0}(v(\mathbf{x})t)$ donc le stress ne change que l'échelle.

Maintenant, considérons le cas classique où les stress ne changent pas au cours du temps. Soient $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{X}$ deux stress, on considère le modèle suivant :

$$R_{\mathbf{x}_2}(t) = R_{\mathbf{x}_1}(\rho(\mathbf{x}_1, \mathbf{x}_2)t),$$

où $\rho(\mathbf{x}_1, \mathbf{x}_2) = F_A(\mathbf{x}_2)/F_A(\mathbf{x}_1)$.

Dans le cas unidimensionnel, le taux de changement d'échelle sous un stress \mathbf{x} est donc :

$$\delta(\mathbf{x}) = \lim_{\Delta \mathbf{x} \rightarrow 0} \left[\frac{\rho(\mathbf{x} + \Delta \mathbf{x}, \mathbf{x}) - \rho(\mathbf{x}, \mathbf{x})}{\Delta \mathbf{x}} \right] = \left(\log(F_A(\mathbf{x})) \right)'$$

Donc :

$$F_A(\mathbf{x}) = \exp \left[\int_{\mathbf{x}_0}^{\mathbf{x}} \delta(x) dx \right].$$

On peut dès lors utiliser les connaissances sur la physique de la dégradation afin de régler le taux de changement d'échelle du modèle accéléré sur la cinétique des équations mécaniques, physique ou chimiques de la défaillance. Si $\delta(\mathbf{x})$ est proportionnelle à une fonction $g(x)$ connue, alors $F_A(\mathbf{x}) = \exp[a_0 + a_1 \cdot g(\mathbf{x})]$.

En supposant que $\delta(\mathbf{x}) = a_1$, on obtient un modèle Log-linéaire :

$$F_A(\mathbf{x}) = \exp[a_0 + a_1 \cdot \mathbf{x}].$$

Si $\delta(\mathbf{x}) = -a_1/\mathbf{x}$, on obtient un modèle de la règle de puissance :

$$F_A(\mathbf{x}) = \exp[a_0 - a_1 \cdot \log(\mathbf{x})] = c \cdot \mathbf{x}^{-a_1}.$$

Et si $\delta(\mathbf{x}) = a_1/\mathbf{x}^2$, on obtient un modèle d'Arrhénius :

$$F_A(\mathbf{x}) = \exp[a_0 + a_1/\mathbf{x}].$$

Dans le cas multidimensionnel, on considère les caractéristiques infinitésimales $\delta_i(\mathbf{x})$ définies comme suit :

$$\delta_i(\mathbf{x}) = \lim_{\Delta \mathbf{x}_i \rightarrow 0} \left[\frac{\rho(\mathbf{x} + \Delta \mathbf{x}_i \cdot e_i, \mathbf{x}) - \rho(\mathbf{x}, \mathbf{x})}{\Delta \mathbf{x}_i} \right] = \frac{\partial \log F_A(\mathbf{x})}{\partial \mathbf{x}_i},$$

où $e_i = (0, \dots, 1, \dots, 0)'$ avec l'unité est à la $i^{\text{ème}}$ composante.

De proche en proche, on trouve finalement une forme générale du facteur d'accélération:

$$F_A(\mathbf{x}) = \exp \left[a_0 + \sum_{i=1}^{r-1} a_i \cdot g_i(\mathbf{x}) \right],$$

où les fonctions $g_i(\cdot)$ sont connues et les paramètres a_0, a_1, \dots, a_{r-1} sont inconnus.

Bagdonavicius, Gerville-Réache, Nikoulina et Nikulin [3] ont étudié en détails le cas paramétrique (cas des lois Weibull, Loglogistique et Lognormale) et le cas semi-paramétrique. De plus, ils ont considéré deux types de plans d'expériences (i) lorsque le stress est constant, (ii) lorsque le stress est une fonction étagée décroissante.

1.8 Modèle AFT

1.8.1 Notations et définitions

Maintenant revenons aux problèmes concernant l'étude de la fiabilité des semi-conducteurs. Nous avons vu dans le chapitre précédent que les principaux modes de défaillances mènent à supposer que la distribution des durées de vie est Lognormale ou de Weibull. Soit $Y = \log(Z)$ le logarithme de la durée de vie Z , et soit \mathbf{x} la covariable (environnementale) associée à Y . Ensuite, posons \mathcal{X} l'ensemble des valeurs des covariables, qui représente pour un modèle accéléré l'ensemble des environnements possibles ou ensemble des stress. Le modèle accéléré est un modèle de régression défini comme suit:

$$Y = f(\mathbf{x}) + \sigma \varepsilon \quad (1.2)$$

où f est une fonction d'accélération (le logarithme de la moyenne pour le cas Lognormal et le logarithme du paramètre d'échelle pour le cas Weibull) définie sur l'ensemble des stress \mathcal{X} , σ^2 est la variance du modèle et ε désigne la variable aléatoire des erreurs. De plus, notons Φ la fonction de distribution des erreurs : pour le cas de durées Lognormales, la distribution Φ est la distribution normale standardisée $\mathcal{N}(0, 1)$ et pour le cas de durées de Weibull, la distribution Φ est la distribution de Gumbel normalisée $SEV(0, 1)$.

Remarque 1.8.1. Remarquons que dans la pratique les composantes de \mathbf{x} sont soit des températures en degré Kelvin, soit une intensité de courant, soit une tension, soit un taux d'humidité ou bien des effets joints.

Dans la plupart des cas pratiques comme on travaille à partir du logarithme d'une durée, il est possible d'écrire le facteur d'accélération de (H2), comme une combinaison linéaire des environnements de stress ou de leurs transformations : on parle alors de modèle AFT (voire livre de Nikulin et al. [28], chap. 4 et 5)

Définition 1.1. Lorsque la *fonction d'accélération* f du modèle d'accélération (1.2) s'écrit comme une combinaison linéaire de \mathbf{x} ou de transformations de \mathbf{x} , on parle alors de modèle AFT pour Accelerated Failure Time. Le modèle AFT généralisé s'écrit comme suit :

$$f(\mathbf{x}) = a_0 + \sum_{j=1}^{r-1} a_j g_j(\mathbf{x}) \quad (1.3)$$

où les $(a_j)_{j=0,1,\dots,r-1}$ sont les paramètres du modèle accéléré et $(g_j)_{j=1,\dots,r-1}$ est une famille finie de transformations de \mathbf{x} issue de la physique de la défaillance lorsque $\mathbf{x} \in \mathcal{X}$. Le plus souvent, les fonctions g_j sont des fonctions polynomiales ou logarithmiques des composantes de \mathbf{x} ou de l'inverse des composantes de \mathbf{x} .

Ensuite les modèles classiques décrits dans la section précédente s'écrivent tous comme un modèle AFT (cf. Mann and al. [23]). Ils sont regroupés dans le tableau (1.4).

Table 1.4: Modèles accélérés classiques.

Nom	Accélération	Covariables
Arrhenius	$f(x) = a_0 + a_1 \cdot \frac{1}{x}$	\mathbf{x} : température
Puissance Inverse	$f(x) = a_0 + a_1 \cdot \log\left(\frac{1}{x}\right)$	\mathbf{x} : tension
Eyring	$f(x) = a_0 + a_1 \cdot \frac{1}{x} + a_2 \log\left(\frac{1}{x}\right)$	\mathbf{x} : température
Generalized Eyring	$f(x) = a_0 + a_1 \cdot \frac{1}{x_1} + a_2 \log\left(\frac{1}{x_1}\right) + a_3 \cdot x_2 + a_4 \cdot \frac{x_2}{x_1}$	$\mathbf{x} = (x_1, x_2)$: x_1 : température x_2 : tension
Black	$f(x_1, x_2) = a_0 + a_1 \cdot \frac{1}{x_1} + a_2 \log\left(\frac{1}{x_2}\right)$	$\mathbf{x} = (x_1, x_2)$: x_1 température x_2 intensité

1.8.2 Estimateur de contraste du minimum

L'estimateur des moindres carrés est l'estimateur naturel du problème de régression linéaire (Rao et al. [30]). Par ailleurs, il est très populaire, comme il s'applique également dans de nombreux cas où la distribution des erreurs n'est pas normale. En effet, il est possible d'utiliser l'estimateur des moindres carrés, lorsque l'on respecte les hypothèses de Gauss-Markov (erreurs centrées et homoscedasticité).

Comme nous l'avons vu, le modèle accéléré généralisé (1.3) se présente comme un modèle de régression linéaire : soit $Y = (y_1, \dots, y_n)$ le vecteur des observations (log-durées), soient $\mathbf{x}_1, \dots, \mathbf{x}_n$ les n covariables associées aux Y_i et soit X la matrice de régression telle que :

$$X = \begin{pmatrix} 1 & g_1(\mathbf{x}_1) & \dots & g_{r-1}(\mathbf{x}_1) \\ 1 & g_1(\mathbf{x}_2) & \dots & g_{r-1}(\mathbf{x}_2) \\ \vdots & \vdots & \ddots & \vdots \\ 1 & g_1(\mathbf{x}_n) & \dots & g_{r-1}(\mathbf{x}_n) \end{pmatrix}$$

L'estimateur des moindres carrés est :

$$\hat{\theta} = (X'X)^{-1}X'Y,$$

où $\hat{\theta}_i$ la composante numéro i de $\hat{\theta}$ est l'estimateur des moindres carrés du paramètre accéléré a_{i-1} de (1.3).

Dans le cadre de modèles accélérés, il peut se trouver limité par deux points :

- un problème de colinéarité des covariables ;
- un problème d'échelle.

En effet, lorsque la fonction d'accélération est complexe, comme la fonction d'Eyring généralisé tab.1.4, il arrive que les colonnes de la matrice de régression X soient liées. Ou bien, lorsque l'échelle de grandeur d'une composante de la covariable est très petite, la valeur propre associée à ce terme est faible et il peut en dériver des erreurs de calculs.

Dans les deux cas, il y aura un problème lors de l'inversion de $(X'.X)$ et donc il y aura un problème lors de l'estimation. Ce genre de problème est bien connu et de nombreuses méthodes permettent de remédier à ces erreurs [30]. Parmi ces méthodes, la régression *ridge* permet de remédier aux problèmes d'échelles ou de colinéarité. Le principe du *ridge* consiste à introduire un terme de régularisation $0 \leq \tau \leq 1$ dans les équations d'estimation et l'estimateur *ridge* est alors :

$$\tilde{\theta} = (X'.X + \tau.I_r)^{-1}.X'.Y,$$

où I_r est la matrice identité de dimension r .

Remarque 1.8.2. Remarquons que pour le cas où la loi de la durée de vie de Z est de Weibull, en effectuant une transformation logarithmique de Z , on retrouve la loi des valeurs extrêmes de Gumbel $SEV(\mu, 1/\sigma)$ de moyenne $(\mu + \sigma.\gamma_0)$ et de variance $\sigma^2.\frac{\pi^2}{6}$, où γ_0 est la constante gamma d'Euler qui vaut approximativement 0.5772. Moyennant une légère transformation des variables aléatoires destinée à vérifier les hypothèses de Gauss-Markov, il est possible de réaliser une approximation linéaire afin d'estimer les paramètres accélérés pour des durées de Weibull.

1.8.3 Maximum de vraisemblance

Le maximum de vraisemblance est un M-estimateur bien connu [35]. Dans le cadre gaussien, on utilise l'estimateur de maximum de vraisemblance (EMV) sur des données censurées. Et dans l'étude de variable aléatoire de Weibull, on utilise communément cet estimateur. Nous présenterons dans ce paragraphe l'EMV pour des données accélérées dans le cadre Weibull.

Comme son nom l'indique le maximum de vraisemblance est une méthode d'estimation basée sur la maximisation de la vraisemblance du modèle. Ainsi contrairement aux moindres carrés, cette méthode est basée sur la résolution d'équations implicites et elle nécessite la plupart du temps l'utilisation d'algorithmes d'approximation tel que la méthode itérative de type Gauss-Newton. Par exemple, pour le cas de données accélérées dans le cadre Weibull, les paramètres β et η de la loi de Weibull sont tels que :

$$\begin{cases} \beta & = 1/\sigma \\ \eta(\mathbf{x}) & = \log \sum_{j=0}^{r-1} a_j g_j(\mathbf{x}) = \log(\mu(\mathbf{x})) \end{cases}$$

avec la convention $g_0(\mathbf{x}) = 1$. Et dans ce cas, la loi de la variable aléatoire du logarithme de la durée de vie est une loi de Gumbel de paramètres (μ, σ) .

La fonction vraisemblance, notée L , du modèle associée aux données accélérées (Y, \mathbf{x}) vaut :

$$L(a, \sigma) = \prod_{i=1}^n \frac{1}{\sigma} \exp \left\{ \frac{y_i - \sum_{j=0}^{r-1} a_j g_j(\mathbf{x}_i)}{\sigma} \right\} \exp \left\{ - \exp \left(\frac{y_i - \sum_{j=0}^{r-1} a_j g_j(\mathbf{x}_i)}{\sigma} \right) \right\}$$

$$\log L(a, \sigma) = -n \log(\sigma) + \frac{1}{\sigma} \sum_{i=1}^n \left(y_i - \sum_{j=0}^{r-1} a_j g_j(\mathbf{x}_i) \right) - \sum_{i=1}^n \exp \left\{ \frac{y_i - \sum_{j=0}^{r-1} a_j g_j(\mathbf{x}_i)}{\sigma} \right\}$$

Dans ce cas, les équations de vraisemblance obtenues en calculant les dérivés partielles de chaque paramètre $(a_0, \dots, a_{r-1}, \sigma)$ sont :

$$\frac{\partial \log L}{\partial a_r}(a, \sigma) = \frac{1}{\sigma} \sum_{i=1}^n g_r(\mathbf{x}_i) \cdot \left[\exp \left\{ \frac{y_i - \sum_{j=0}^{r-1} a_j g_j(\mathbf{x}_i)}{\sigma} \right\} - 1 \right]$$

$$\frac{\partial \log L}{\partial \sigma}(a, \sigma) = -\frac{n}{\sigma} + \frac{1}{\sigma^2} \sum_{i=1}^n \left(y_i - \sum_{j=0}^{r-1} a_j g_j(\mathbf{x}_i) \right) \cdot \left[\exp \left\{ \frac{y_i - \sum_{j=0}^{r-1} a_j g_j(\mathbf{x}_i)}{\sigma} \right\} - 1 \right]$$

L'EMV $(\hat{a}, \hat{\sigma})$ est obtenu en résolvant le système d'équations implicites ci-dessus en zéro. Les propriétés de l'EMV sont bien connues : estimateur consistant, normalité asymptotique et efficacité asymptotique. En pratique, il est important de se donner un indicateur permettant de préciser cette notion asymptotique. Mann et al. [23] proposent d'étudier la forme de la fonction de vraisemblance relative afin de constater la normalité ou la non normalité de l'EMV. L'objet de la fonction de vraisemblance relative est de regarder composante par composante le rapport maximal entre la fonction de vraisemblance sous la contrainte qu'une composante est fixée et la fonction de vraisemblance de l'EMV. Pour la première composante de a_0 la fonction de vraisemblance relative \mathcal{R}_m est définie telle que :

$$\mathcal{R}_m(a_0) = \max_{(a_1, \dots, a_{r-1}, \sigma)} \frac{L(a_0, a_1, \dots, a_{r-1}, \sigma | Y, \mathbf{x})}{L(\hat{a}, \hat{\sigma} | Y, \mathbf{x})}.$$

Lorsque la fonction de vraisemblance relative est symétrique alors la loi de l'EMV est approximativement Gaussienne. En effet, regarder les fonctions de vraisemblance relative pour chaque composante revient à étudier le rapport de comportement du rapport de vraisemblance entre le « *vrai* » modèle et le modèle estimé. En vertu du troisième lemme de Le Cam, on peut déduire de la normalité du rapport de vraisemblance la normalité des estimateurs.

1.8.4 Application et Arrhénius Lognormale

Dans ce paragraphe, nous allons illustrer l'estimation de modèle accéléré pour un jeu de données de loi Lognormale, puis nous parlerons du plan d'expérience. Et ensuite nous estimerons un modèle accéléré Lognormal.

Nous avons utilisé des données du livre de Nelson [27] afin d'illustrer une application du modèle Lognormal d'Arrhénius. Les données de l'exemple sont des moteurs électroniques dont on souhaite étudier la fiabilité lorsque la température augmente. Le modèle d'accélération classique pour des composants sujet à un stress thermique est le modèle d'Arrhénius et on suppose que la loi de distribution des durées de vie est Lognormale. Soient $Y = (y_1, \dots, y_n)$ les réalisations d'un n -échantillon du logarithme des durées de vie et $\mathbf{x} = (\mathbf{x}_1, \dots, \mathbf{x}_n)$ les covariables associées aux observations Y . Pour le cas du modèle d'Arrhénius, $\mathbf{x}_i = \frac{1}{\kappa \cdot T_i}$ est l'inverse de la température T_i exprimée en degré Kelvin multipliée par la constante de Boltzmann $\kappa = 8.6171 \cdot 10^{-5} eV/K$. Le modèle accéléré est alors :

$$\begin{aligned} y_i &= f(\mathbf{x}_i) + \sigma \varepsilon_i, \\ \text{et} \\ f(\mathbf{x}_i) &= a_0 + a_1 \mathbf{x}_i; \end{aligned}$$

où les paramètres (a_0, a_1) définissent la relation d'accélération, σ est la paramètre de forme des durées de vie et les (ε_i) sont des variables aléatoires des erreurs *i.i.d.* et dont la loi jointe est une Gaussienne centrée et réduite.

Le plan d'expérience de cet essai a été réalisé selon une méthode traditionnelle (mais non optimale). On choisit N différents niveaux de stress utilisés durant l'essai. Et on répartit régulièrement ces stress entre une valeur minimale de stress $\mathbf{x}_M = \frac{1}{\kappa \cdot (190 + 273.15)}$ et une valeur maximale de stress $\mathbf{x}_H = \frac{1}{\kappa \cdot (260 + 273.15)}$; ces deux valeurs sont choisies par avis d'expert ou selon les standards. Ensuite, on répartit régulièrement les n unités tests entre les N environnements de test.

Cette méthode traditionnelle de construction du plan d'expérience est basée sur la variance de l'estimateur de $\mu(x_0)$ où \mathbf{x}_0 désigne le niveau de stress en condition opérationnelle. Pour les modèles log-linéaires Lognormaux, la variance de $\hat{\mu}(\mathbf{x}_0)$ vaut [27] :

$$Var(\hat{\mu}(\mathbf{x}_0)) = \left[1 + (\mathbf{x}_0 - \bar{\mathbf{x}})^2 \frac{n}{(\sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})^2)} \right] \frac{\sigma^2}{n};$$

où $\bar{\mathbf{x}} = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i$. Ainsi avec $N = 4$ et une température en condition opérationnelle de 180 degrés Celsius ; la variance du modèle vaut $[1 + \frac{36}{5}(1.17 - 0.5)^2] \frac{\sigma^2}{n} = 4.23 \frac{\sigma^2}{n}$. Ensuite n la taille de l'échantillon est choisie en fonction de cette variance et de la précision souhaitée. Ainsi en utilisant une approximation normale, en se donnant une marge d'erreur de ± 0.1 et en utilisant $s(0) = .01053$ une estimation issue du retour d'expérience de σ , alors $n \geq [1 + 7.2(1.17 - 0.5)^2] \frac{(1.96s)^2}{0.1^2} \approx 19$; donc il faudrait mettre au moins 5 unités de test par environnement de stress. Dans le présent essai, on a placé 10 unités tests, soit $n = 40$.

En réalisant une régression simple sur les données issues de l'essai accéléré, on note que les résidus ne sont pas aléatoirement répartis. La variance augmente avec le niveau de stress. Nous avons ainsi estimé l'écart type de chaque niveau de stress et puis nous avons supposé que la variance de ε est $V = (v_{i,j})_{1 \leq i,j \leq n}$, telle que $v_{i,j} = 0$ si $i \neq j$ et $v_{i,i}$ vaut la variance estimée sous le stress \mathbf{x}_i . On estime alors par les moindres carrés normalisés : $\hat{a}_0 = -3,6234$; $\hat{a}_1 = 3,4956$ et $\hat{\sigma} = 1,065$.

La figure 1.8.4 illustre la qualité de la régression. La figure regroupe à gauche le nuage des points ainsi que la droite de régression estimée. Et le graphique de droite montre la répartition empirique des résidus en échelle logarithmique ; on voit ainsi que la loi des log-durées observées est bien normale.

Remarque 1.8.3. Le plan d'expérience traditionnel n'est pas optimal. Deux méthodes sont souvent recommandées [27] :

1. Tester deux environnements de stress \mathbf{x}_H et \mathbf{x}_M ; puis déterminer p la proportion des unités tests soumis au stress \mathbf{x}_H . La proportion p^* minimisant la variance de $\hat{\mu}(\mathbf{x}_0)$ est $p^* = \frac{\xi_0}{(2\xi_0-1)}$ où $\xi_0 = \frac{\mathbf{x}_H - \mathbf{x}_0}{\mathbf{x}_H - \mathbf{x}_M}$;
2. Tester trois ou quatre environnements de stress régulièrement répartis entre \mathbf{x}_H et \mathbf{x}_M ; puis comme en 1. proposer d'augmenter le nombre d'observations des plus bas niveaux de stress. Ce type de test est plus complexe, mais il permet de mettre en avant des phénomènes nouveaux tels que le surstress (voir section 1.10).

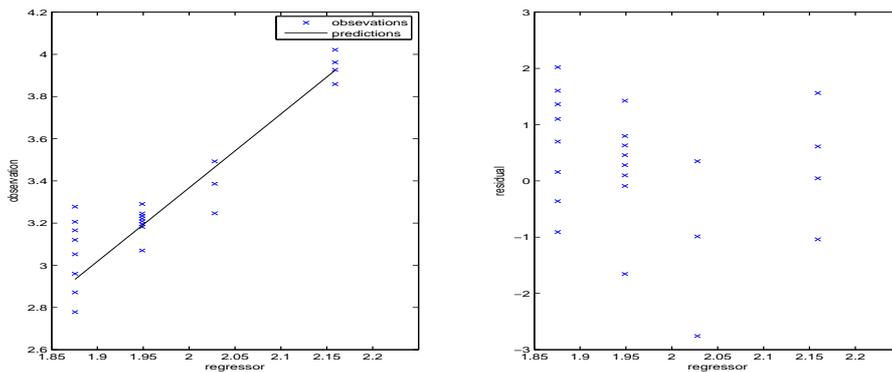


Figure 1.15: Exemple d'essai accéléré Arrhénus Lognormal estimé par les moindres carrés: en a) on retrouve le nuage de points et la droite de régression ; et en b) le graphique des résidus.

1.8.5 Application et modèle de Weibull

Dans ce paragraphe, nous allons illustrer l'estimation de modèle accéléré pour un jeu de données de loi de Weibull. Nous avons repris un échantillon du livre de Nelson [27]. Il s'agit d'un essai accéléré de 76 composants électroniques dont on souhaite tester l'isolation. Ainsi la variable de stress est la tension (proportionnel à l'intensité de courant selon la loi d'Ohm). On applique ici une relation de puissance. La tension varie de 26 à 36 kV, le nombre d'observations par niveau de stress varie (5 à 26kV, 7 à 28 kV, 11 à 30 kV, 15 à 32 kV, 19 à 34 kV, 15 à 36 kV et 8 à 38 kV) et on mesure le temps avant court-circuit en heures (ici il n'y a pas de censure).

Pour cet essai, on applique la relation d'accélération de puissance inverse. Notons Y le logarithme des durées de vie et \mathbf{x} le logarithme de l'inverse des tensions appliquées. Le modèle accéléré est défini de telle sorte que :

$$Y = a_0 + a_1 \cdot \mathbf{x} + \sigma \cdot \varepsilon;$$

où (a_0, a_1, σ) sont les paramètres à estimer et ε est un variable aléatoire des erreurs. Ici la loi de distribution des erreurs est une loi de Gumbel $SEV(0; 1)$. En vue d'utiliser l'estimateur des moindres carrés, ici notons que la loi de Gumbel n'est ni centrée, ni réduite. L'espérance d'une loi $SEV(0; 1)$ est la constante $\gamma \approx 0.5772$ de Euler, et son écart type est $\frac{\pi}{\sqrt{6}}$. En tenant compte de ces transformations, on obtient par la méthode des moindres carrés que :

$$\hat{a}_{0,MC} = 60,8482 ; \hat{a}_{1,MC} = 16,3909 \text{ et } \hat{\sigma}_{MC} = 1,5582.$$

Au regard de l'estimation et de la dispersion des résidus montrés en 1.8.5, l'analyse des données issues des essais accélérés ne pose aucun problème particulier. Notons que si on met en place l'estimateur du maximum de vraisemblance (ici en utilisant la méthode itérative de Levenberg-Marquardt), on retrouve des résultats comparables (et ici rigoureusement identiques pour α_1 le paramètre ayant le plus de sens physique):

$$\hat{a}_{0,MV} = 59,4468 ; \hat{a}_{1,MV} = 16,3909 \text{ et } \hat{\sigma}_{MV} = 1,5375.$$

Notons par ailleurs qu'en appliquant un bootstrap paramétrique de l'EMV sur 300 ré échantillonnages de (Y, \mathbf{x}) . La figure 1.17 résume ce bootstrap et elle présente la distribution bootstrap des paramètres de l'EMV. On remarque que les EMC sont toujours dans l'intervalle de confiance bootstrap des EMV.

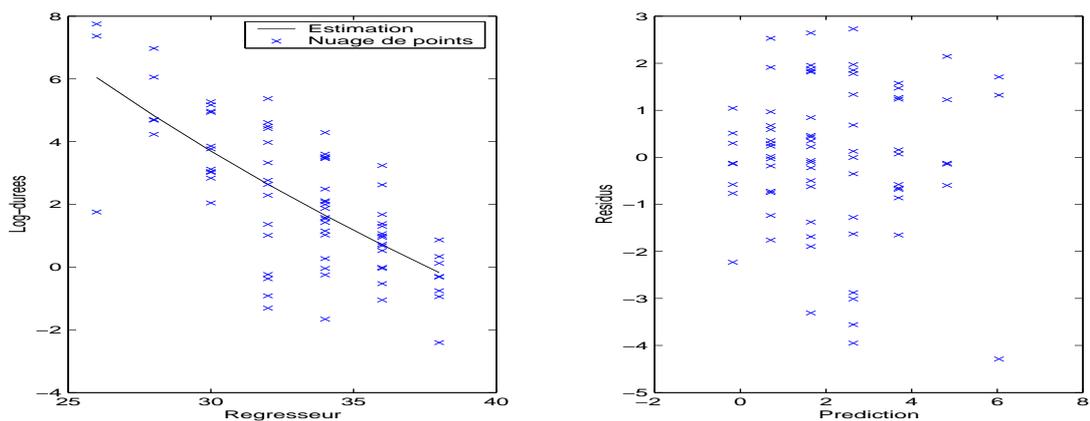


Figure 1.16: Exemple d'essai accéléré puissance-Weibull estimé par les moindres carrés : en a) on retrouve le nuage de points et la droite de régression ; et en b) le graphique des résidus.

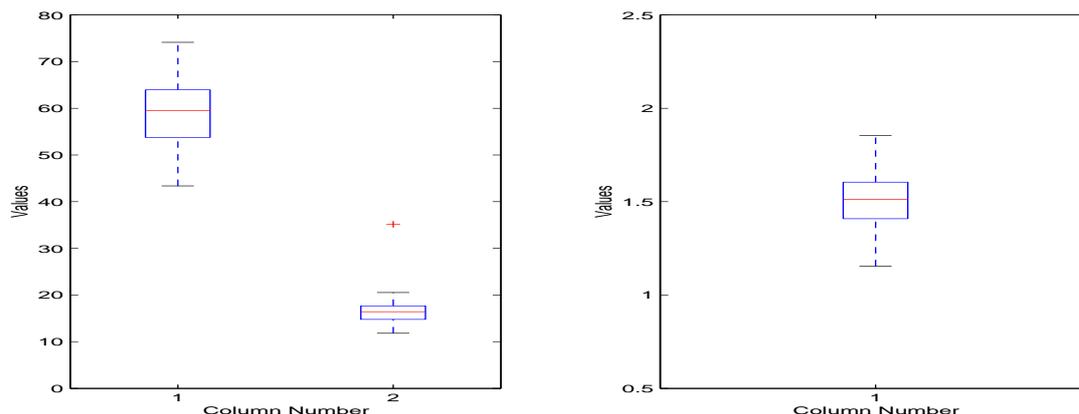


Figure 1.17: Exemple d'essai accéléré puissance-Weibull estimé par le maximum de vraisemblance, résultats d'un bootstrap paramétrique de l'EMV sur 300 répétitions : en a) on retrouve le boxplot des paramètres $\hat{a}_{0,MV}$ et $\hat{a}_{1,MV}$; et en b) le boxplot de $\hat{\sigma}_{MV}$.

1.9 Limites des hypothèses d'accélération

1.9.1 Généralités

Lorsque les connaissances du mécanisme de défaillance sont en accord avec les hypothèses du modèle d'accélération (1.1) ; alors les données recueillies lors d'un test ALT peuvent être analysées par les modèles accélérés classiques, voir Mann et al. ([23], 1974), Nelson (1990, [27] ou Meeker et Escobar ([25], 1998). Toutefois, lorsqu'il arrive que les hypothèses (1.1) ne sont pas respectées, les caractéristiques de fiabilité estimés à partir d'un tel test ALT sont clairement erronés. En vue d'étudier les limites des

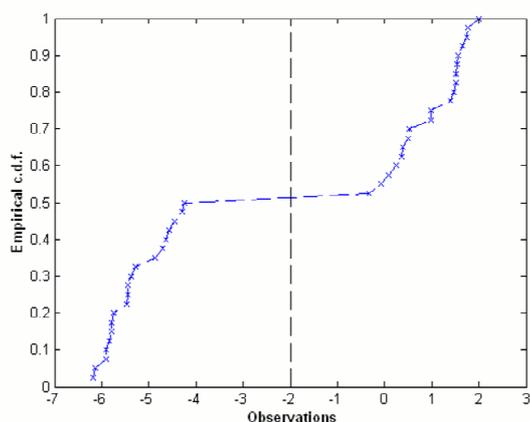


Figure 1.18: Distribution de données accélérées lorsqu'il y a deux plages d'accélération masquées, on observe un « plateau » représentant le changement de mode.

hypothèses d'accélération, Meeker et Escobar ([25]) ont dressé une liste exhaustive de situations dans lesquelles il est nécessaire d'adapter de nouvelles hypothèses aux données accélérées, comme par exemple :

- Le paramètre de forme σ évoluant selon une fonction d'accélération ;
- Modes de défaillance multiples et inconnus ;
- Fonction d'accélération inadaptée ;
- Plusieurs plages d'accélération masquées ;
- ou autres ; ...

Dans le domaine des puces microélectroniques, nous avons constaté récemment l'apparition de tels phénomènes et notamment l'apparition de plusieurs plages d'accélération. Par exemple, Ogawa et al. [29] ont étudié ce type de phénomène lors d'étude d'électromigration sur les interconnexions construites à partir du procédé de double damascène (optimisant le dépôt des interconnexions sur deux niveaux). Au cours de leur étude, les auteurs ont montré qu'en utilisant les modèles traditionnels on omet un phénomène d'échelle de stress. Ils ont repéré qu'aux bas niveaux de stress, l'électromigration est localisée au dessus de la jonction et qu'elle se propage le long de la ligne de métal. Et aux hauts niveaux de stress, l'électromigration engendre un trou au niveau d'une jonction métallique. Lorsque l'on trace la distribution, un plateau apparaît clairement. Ainsi pour les niveaux de stress choisis par le plan d'expérience, on a révélé deux phénomènes d'accélération.

Remarque 1.9.1. Ce type de comportement montre que lorsqu'on optimise le plan d'expérience, il est important d'avoir plus de deux niveaux de stress ; afin de repérer ce comportement atypique.

1.9.2 Multiples plages d'accélération

Dans le domaine des semi-conducteurs modernes, Ogawa et al. ([29]) ont mis en évidence des problèmes lorsque l'on met en place un test accéléré classique pour étudier la fiabilité d'interconnexions à haut niveau d'intégration, à savoir de l'ordre du dixième de micron. On observe généralement deux types de comportements : soit des risques compétitifs, soit plusieurs plages d'accélération lors d'un test ALT. Dans le cas des risques compétitifs, le taux de défaillance h du modèle est la somme de plusieurs taux de défaillance. Nous laisserons de côté cette problématique, pour plus d'informations voire par exemple le chapitre 7 du livre de Nelson [27]. Et lorsque l'on observe plusieurs plages d'accélération, on parle alors de *surmenage* ou de *sur-stress*. En microélectronique, un défaut est lié à un grand nombre de composantes telles que : l'architecture de la structure, la nature des matériaux, les procédés de dépôts de matières, les couches interfaces entre éléments isolants et conducteurs, l'échelle d'intégration, etc ... Ainsi, à la faveur d'un changement technologique ou lors d'un changement de processus d'intégration, on peut observer que les limites de tolérance du composant changent : on voit alors apparaître plusieurs plages d'accélération, illustré par la figure 1.19. D'autre part,

on peut volontairement surcharger l'essai afin de rechercher les limites de tolérance du composant test. De même on voit alors apparaître plusieurs plages d'accélération. Ce genre de phénomène est bien connu dans d'autres domaines d'application comme les industries plastiques.

Nous proposons donc dans les prochaines parties de ce document de traiter le problème de sur stress dans les deux cadres standards de test accéléré : dans un premier temps, nous considérerons la loi de durée de vie suit une loi Lognormale ; et ensuite, nous supposerons que la loi de durée de vie suit une loi de Weibull.

Remarque 1.9.2. Sur la figure 1.19, on voit un modèle accéléré avec deux plages d'accélérations. Ici, la fonction d'accélération classique est remplacée par une fonction d'accélération définie par morceaux. Notons que ces deux modèles ne sont pas en concurrence, comme ils s'adaptent chacun à des conditions expérimentales précises comme stipulé dans l'hypothèse (H2).

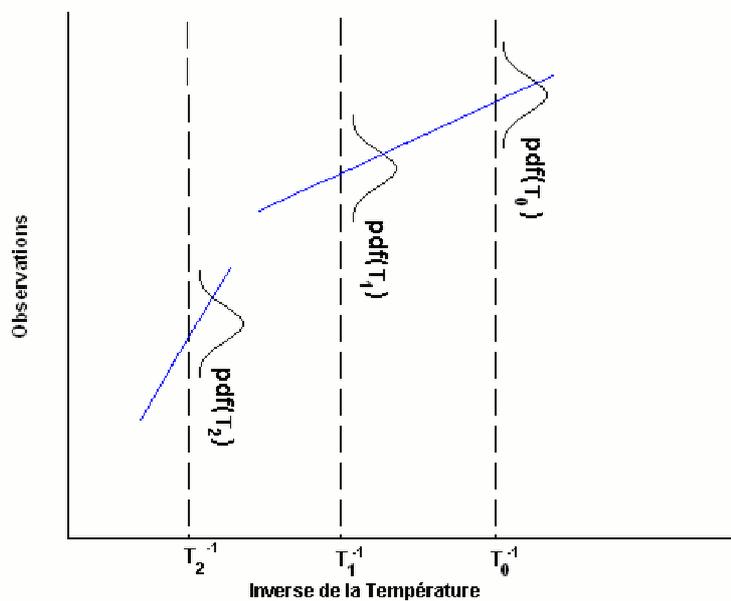


Figure 1.19: Exemple d'un modèle d'accélération à plusieurs plages d'accélérations.

Chapter 2

Model Selection for Piecewise Regression Models and Detection of Overstress on Lognormal Data

Cette partie a été écrite sur la base de deux communications réalisées en collaboration avec Patrick Pamphile : une première communication pour le congrès Qualita de 2005 [17] et une seconde communication pour le congrès Lambda-Mu de 2006 [18].

Abstract : Innovations in technologies commonly leads to met overstress during reliability studies. Then acceleration model usually introduces heavy misleading on the models. When data are subject to overstress, we propose to get piecewise acceleration model. In this work, we have proposed to use a model selection method in order to estimate the number of stress sets. We have implemented the criterion of Birgé and Massart and we have used heuristic calibration in order to build a data-driven criterion. Lastly we have compared our method with classical ones (AIC, AIC_c or BIC) for reliability on microelectronic fields.

Keywords : Acceleration Life Testing, AFT model, Piecewise Regression, Gaussian Model Selection, Penalized Criteria, Overstress Accelerating Data.

Résumé : En fiabilité, le phénomène de sur-stress est couramment observé lors de changements technologiques majeurs. Les modèles accélérés usuels peuvent alors nous induire gravement en erreur. Lorsque les données accélérées sont sujettes au sur-stress nous proposons d'utiliser un modèle d'accélération par intervalle de stress. Dans ce travail, nous proposons d'utiliser une méthode de sélection de modèles afin d'estimer le nombre d'intervalles de stress. Nous proposons d'appliquer le critère de Birgé et Massart en utilisant des heuristiques de calibration afin de construire une méthode de sélection adaptée aux jeux de données. Enfin nous avons comparé cette méthode aux critères classiques (AIC, AIC_c et BIC) dans le cadre de l'étude de fiabilité en microélectronique.

Mots-Clés : Test Accéléré, modèle AFT, Régression par morceaux, Sélection de Modèle Gaussien, Critère Pénalisé, Données Accélérées en Sur-stress.

2.1 Introduction

2.1.1 Generality about ALT

A vast number of sophisticated manufacturing processes are required to complete a sub micron interconnection device. Its failure mechanism is therefore complex and variable. Each failure may have many causes: contamination of the silicon wafer, improper etching of oxide film, corrosion ... ; and it depends on various factors: design, manufacturing, or operating environmental factor. Quality and reliability improvement have to go through statistical analysis of the failure time. Statistic life tests are performed by manufacturers to obtain reliability information of an existing device, or to compare manufacturing processes during the development of a new device. Microelectronic manufacturers widely use Accelerated life test (ALT) [27], to get reliability information in the shortest development time. ALT consists of placing the test units in higher than usual stress environment. That kind of experiment shortly produces degradations without changing the mode of failure. Thus, statistical inference from «accelerated testing» gathered from ALT yields reliability information on devices in actual environment.

Analysis of ALT data consists of making statistical inference in order to completely identify:

1. the lifetime distribution; specified by the failure mode;
2. the « *life-stress relationship* »; specified by the physics of failure.

The lifetime distribution is closely link to the mode of failure involved: Electromigration, Stressmigration, *etc.*. And the life-stress relationship is also called « *acceleration function* » that quantifies how reliability information varies with the stress. It comes from the study of the physical, chemical or mechanical phenomena, or their joined effect, which describe the physical degradation. The Log-Location Scale model is generally used. For interconnections' reliability, this model is a simple regression of the log lifetimes. After, the choice of an acceleration model comes from expert advices and physics of failure [24, 29]. The life-stress relationship is chosen in order to mimic the kinetic of the degradation mechanism. In practice, we use the classical acceleration models as Arrhenius, Power Rule, Eyring or Generalized Eyring relationship [23, 24, 27].

During this work, we will exclusively talk about constant stress ALT, when the stress levels are fixed during the experiment.

2.1.2 Why introducing a selection procedure ?

In the last ten years, the technical evolution in microelectronics' industry has been considerable: inter-layer have become less and less conductor, aluminium interconnects have been replaced by copper or copper with alloy, new interconnection's design, called dual-damascenes, have been integrated. For those chips resulting from deep sub micron

technologies the statistical models traditionally lead to a clear disagreement with the data [29]. Indeed, for those new chips, failure mechanisms are becoming highly complex and partially unknown. Moreover, even if the acceleration model is well-known, its use of range has changed. That is why we propose here to consider a piecewise Log-Location Scale model. This model matches all classical models; it allows overstress tests and is more robust to multi failure mechanisms. However, it is well-known that too large statistical models can produce unstable estimations. That is why we propose to use some selection method in order to fit our model to the data.

Selecting a suitable model from a large class of plausible models is an important problem in statistics. Several criteria are used for model selection, among them Akaike's AIC [1] and Schwarz's BIC [33] are widely accepted. An inherent problem is that their definitions rely heavily on asymptotic considerations. In [6], the authors propose a penalized criterion free of asymptotic definitions. And thanks to this new penalized criterion and a data-driven procedure [6, 21, 19]; we have implemented their approach here.

2.2 Piecewise acceleration model

Many authors have discussed about ALT with parametric statistical methods, as Mann, Shafer and Singpurwalla [23], Nelson [27] or Meeker and Escobar [26]. Here let us talk about ALT practice to sub micron interconnection's reliability. At first, we define the lifetime distribution and the acceleration model. After, we propose a piecewise acceleration model in order to deal with technical innovations.

2.2.1 Lifetime distribution

Let Z be the random variable of the lifetimes, and let us define Y as the random variable of the log lifetimes $Y = \log(Z)$. We assume that:

$$Y = \mu + \sigma \varepsilon, \quad (2.1)$$

where μ is a location or a medium tendency parameter, σ is a scale or a dispersion parameter and ε is the error random variable.

The distribution of ε depends on the failure mode revealing by ALT. Thanks to physical studies on interconnection's reliability [24, 29]; we know that the main failure mode is metal creeping induced by a slow material migration on the device. That is why we assume that the failure mode is a classical aging phenomenon and the distribution of the errors ε is a standardized Gaussian distribution.

2.2.2 Accelerated Failure Time model

Let x be the stress covariables. Let us notice that it can be made of r single stress covariable; as temperature, voltage, current density *etc.*. The Accelerated Failure Time model, denoted AFT model [28], is usually defined through the following life-stress relationship, denoting f , and links the medium tendency parameter μ with the stress \mathbf{x} :

$$\mu = f(\mathbf{x}) = a_0 + \sum_{j=1}^{r-1} a_j g_j(\mathbf{x}), \quad (2.2)$$

where the family of function $(g_j)_{1 \leq j \leq r-1}$ are known by expert advices or the physics of failure. They are widely chosen in order to mimic the kinetic of the degradation mechanism and the relationship depends on the stress \mathbf{x} .

For interconnection field, thanks to Mc Pherson and Dunn [24] and Ogawa and al. [29], who have exhibited diffusion equation of the metal creeping. The three acceleration models, associated with interconnection's reliability, are the two following ones:

Name	Acceleration	Covariables
Arrhenius	$f(x) = a_0 + a_1 \cdot \frac{1}{x}$	\mathbf{x} : temperature
Eyring	$f(x) = a_0 + a_1 \cdot \frac{1}{x} + a_2 \log\left(\frac{1}{x}\right)$	\mathbf{x} : temperature
Black	$f(x_1, x_2) = a_0 + a_1 \cdot \frac{1}{x_1} + a_2 \log\left(\frac{1}{x_2}\right)$	$\mathbf{x} = (x_1, x_2)$: x_1 temperature x_2 density of current

Remark 2.2.1. Let us notice that for Eyring's acceleration model can lead to collinear problem or calculation problem during the estimation. Because the range of the terms $\frac{1}{x_1}$ and $\log\left(\frac{1}{x_2}\right)$ may be very different. We usually take care of it by changing the unity of the current from Ampere to milliamperere or by factorizing the temperature by Boltzmann constant. We also could use some robust estimator, as an example replacing Gaussian regression by a ridge regression.

2.2.3 Limits of the model

Since the thermodynamic theory, we know that the properties of materials depend on the environmental conditions. And the acceleration model is also valid for a specific stress set. When the level of stress is too high, the ALT reveal one (or more) overstress(es) default(s). Such a phenomenon is called overstress. If the model does not take overstress, the ALT would lead to serious disagreement with the reliability yield on the operating conditions. In practice we often encounter overstress when we realize ALT after some technical innovation.

Let us notice that another pitfall of ALT : the competing failure modes [27]. However, on semiconductor's field, we rarely observe competing failure modes, but we may observe overstress fail. Thus the ALT are implementing for very specific structures (also called critical structures) which have been chosen in order to involve the reliability

of the chip. As the test units are much localized. That is why we rarely observe competing failure modes. But when one is incoming, one fail is associating with the lower levels of stress (and the operating stress level) ; and an other one is associating with higher stress levels. But at the medium stress levels, the two fails may occur. At those medium stress conditions, there is competing failure modes.

2.2.4 Piecewise model

We propose a piecewise acceleration model; such a model is natural and adapting to overstress data. We define a piecewise acceleration model as follows:

$$f(\mathbf{x}) = \sum_{J \in M} \left[a_{0,J} + \sum_{j=1}^{r-1} a_{j,J} g_j(\mathbf{x}) \right] \cdot \mathbb{I}_{(x \in J)}, \quad (2.3)$$

where M is a distinct partition of the stress set \mathcal{X} . Every disjoint subset J of M is associated with an acceleration phenomenon acting on the test unit at a stress level J .

This piecewise acceleration model is a natural statistic for ALT when the levels of stress are misunderstood. Furthermore the piecewise acceleration model is also a simple and pertinent model when the acceleration is unknown or when the data reveal two or more failure modes. Each acceleration set is associating with a specific level of stress and a specific location. Moreover, we believe that the piecewise acceleration model approaches the stress levels when the stress condition involves overstress. And when the acceleration reveals some competition failure mode between two accelerations levels ; we shall see two or three stress sets : first the lower stress failure, the next the lower and higher failure on competition and at last the higher failure mode (which may be invisible when the range of the test is not large enough).

2.3 Estimating the number of accelerations

2.3.1 Estimation of the piecewise model

Let M be a partition of \mathcal{X} , with k_M distinct subsets: $\mathcal{X} = \bigcup_{J \in M} J$. And let \mathcal{F}_M denote the set of piecewise acceleration functions on M :

$$\mathcal{F}_M = \left\{ f : \mathbf{x} \longrightarrow f(\mathbf{x}) = \sum_{J \in M} \left[a_{0,J} + \sum_{j=1}^{r-1} a_{j,J} g_j(\mathbf{x}) \right] \cdot \mathbb{I}_{(x \in J)} : \text{where } a_{j,J} \in \mathbb{R} \right\} \quad (2.4)$$

Consider that (Y, \mathbf{x}) is an accelerated framework of (2.1) and (2.3), where $Y = (Y_1, \dots, Y_n)$ is a n -sample of observations, and $\mathbf{x} = (\mathbf{x}_1, \dots, \mathbf{x}_n)$ is the stress vector associated with Y . Now the estimating problem is a simple regression problem. Let \hat{f}_M denote the least squares Estimation (LSE) of f on acceleration functions \mathcal{F}_M .

Larger is k_M , better is the fit and smaller is the approximation error. But an estimation with large k_M can be unstable with large estimation error. A selection model allows reaching a compromise between these two terms.

2.3.2 Penalized least squares Estimator

Let M_0 be a given disjoint partition of \mathcal{X} . And defining \mathcal{M}_n as the finite set of all disjoint partitions of \mathcal{X} build by using the partition M_0 as a canonical basis. We consider the collection of estimator $\{\hat{f}_M : M \in \mathcal{M}_n\}$ and we have to chose the «best» estimator by using a penalized criterion.

Let pen be a penalized function from \mathcal{M}_n to $\mathbb{R}+$ and increasing with k_M . The penalized criterion for M is defined as follows:

$$\begin{aligned} crit(M) &= \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{f}_M(\mathbf{x}_i))^2 + pen(M), \\ &= \|Y - \hat{f}_M\|_n^2 + pen(M), \end{aligned} \quad (2.5)$$

where the norm $\|\cdot\|_n$ is the Euclidian norm of \mathbb{R}^n factorized by $n^{-\frac{1}{2}}$. Moreover, we have considered here function \hat{f}_M as some vector of \mathbb{R}^n composed namely :

$$(\hat{f}_M(\mathbf{x}_1), \dots, \hat{f}_M(\mathbf{x}_n)).$$

And the penalized estimator is $\tilde{f} = \hat{f}_{\hat{M}}$, where :

$$\hat{M} = \arg \min_{M \in \mathcal{M}_n} crit(M). \quad (2.6)$$

The first term of the criterion indicates the goodness of fit: the larger k_M is, the better the fit. On the other hand, the penalty's term allows controlling the complexity of the model: the larger k_M , the higher the variance of the estimation. Finally the penalized criterion attempts to get a compromise between the bias term and the variance term.

In the literature, many penalized criteria have been proposed, as Akaike's Information Criterion [1] or Bayesian's Information criterion [33]. Both criteria come from asymptotic considerations on the risk estimation for AIC and on the conditional risk estimation for BIC. When n is small, Hurvish and Tsai [15] have proposed a correction on AIC (AICc).

For these classical criteria and for the model used in ALT and defined in (2.1) and (2.3), the penalties functions are :

Method	Penalty function
AIC	$pen(M) = 2.r.k_M.\sigma^2$
BIC	$pen(M) = r.k_M.\log(n).\sigma^2$
AICc	$pen(M) = \frac{(n+r.k_M)}{(n-r.k_M-2)}.\sigma^2$

The classical penalized criteria are based on the study of the average on the loss function of the estimator (here the least squares estimation). However, when the number of models increases, the range of the average of $\|Y - \hat{f}_M\|_n^2$ changes. In such a case, those classical criteria may lead to large overestimation (referring to the number of parameters of the model). In order to get free from this misleading, we usually study

the concentration of the $\|Y - \hat{f}_M\|_n^2$ over its expectation by using some concentration inequalities (see works of Talagrand and Ledoux). When the framework are Gaussian regression data Baron, Birgé and Massart [5] have built a new penalized criterion by studying the concentration of $\|Y - \hat{f}_M\|_n^2$ over its expectation.

2.3.3 Selection theorem

In a general study, Baron, Birgé and Massart [5] have proposed a method that leads to get a general expression of the penalty function by using new concentration inequality. Birgé and Massart [6, 7] have also discussed about the Gaussian model selection. They have proposed a selection theorem and a general expression of the penalty function. In our case, the selection theorem leads to the following proposition :

Theorem 2.1. *Suppose that σ is known. Let $\{w_M\}_{M \in \mathcal{M}_n}$ be a family of weights defined as follows:*

$$\sum_{M \in \mathcal{M}_n} \exp(-w_M) = \Sigma < +\infty \quad (2.7)$$

Let $K > 1$ and assume that:

$$\text{pen}(M) \geq K \cdot \frac{\sigma^2}{n} \left(\sqrt{r \cdot k_M} + \sqrt{2 \cdot w_M} \right)^2. \quad (2.8)$$

Then the penalized least squares estimator \tilde{f} exists and it is unique. Furthermore, the following inequality holds:

$$\mathbb{E}_f \left[\|f - \tilde{f}\|_n^2 \right] \leq C(K) \cdot \left\{ \inf_{M \in \mathcal{M}_n} \left(\|f - f_M\|_n^2 + \text{pen}(M) \right) + (1 + \Sigma) \cdot \frac{\sigma^2}{n} \right\} \quad (2.9)$$

where $C(K)$ depends only on K and f_M denotes the orthogonal projection of f on \mathcal{F}_M with respect to the norm $\|\cdot\|_n$.

Remark 2.3.1. Nowadays many papers have developed the works of Baron and al. [5] for several kind of models : see as an example for regular histograms see the work of Birgé and Rozenholc [8]; for Gaussian models refer to Birgé and Massart [6, 7], Massart [21], or Lebarbier [19]; for regression models with sub gaussian data refer to the works of Baraud, Comte and Viennet [4]; for piecewise polynomial models with non gaussian data see the works of Castellan [11] or Sauvé [31].

The previous theorem characterizes the penalty function that leads to a «good» criterion with some constant K and some family of weights $(w_M)_{M \in \mathcal{M}}$. The quality of the penalized criterion defined by (2.8) will be measured thanks to the risk bound (2.9). Let us notice that it depends on K and $(w_M)_{M \in \mathcal{M}_n}$ via pen and Σ . Now we have to discuss about the meaning of those terms and also talk about how to get them.

The constant K is a theoretical constant and it comes from the several bound used to get the risk bound (2.9). That is why we see a term $C(K)$ which only depends on K . In a previous paper, Birgé and Massart [6, 7] have proved that the range of the

value of K shall be large enough to assure the consistency of the penalized estimator \tilde{f} . So, in practice we have to find K as it is large enough to induce the consistency of \tilde{f} , and not too large in order to keep $C(K)$ close to 1.

Furthermore, the constant K is usually calibrating to the data [7]. On the following sections we will discuss of heuristics of calibration. Otherwise, the weight family $(w_M)_{M \in \mathcal{M}}$ are usually interpreting as a way to define the *a priori* probability law of the models \mathcal{F}_M when $M \in \mathcal{M}_n$. In the next section, we will talk about the choice the weight family $(w_M)_{M \in \mathcal{M}}$.

2.3.4 Form of the penalty function (choice of the weights)

The role of the weights can be interpreted as an implement which reflects the complexity of the models when the dimension increases. They clearly define the form of the penalty function. Two strategies are usually made to choose their values: the simplest way is to choose constant weights, $w_M = w > 0$; and, the other way is to choose weights w_M varying with the dimension.

Constant weights

Let us recall that the weights have to respect the suitability condition of the exponential of the weights. Let us denoting N_n as the largest number of acceleration, which is also the number of pieces of the fine partition M_0 . If we take care of the definition of Σ , then the following inequality holds:

$$\begin{aligned} \Sigma &= \sum_{m \in \mathcal{M}} e^{-w} \\ &= \sum_{k=1}^{N_n} \#\{M \in \mathcal{M}_n : k_m = k\} \cdot e^{-w} \\ &\leq \sum_{k=1}^{N_n} \binom{N_n - 1}{k - 1} e^{-w} \\ &\leq (2^{N_n - 1} - 1) \cdot e^{-w}. \end{aligned}$$

Thus the selection theorem gives us the general form of the penalty (2.8), we are able to get the form of *pen*. When the weights are constants, the general form of the penalty function is:

$$pen(M) = c.r.k_M \cdot \frac{\sigma^2}{n} \quad (2.10)$$

where the calibration constant c depends on Σ and r . Let us notice that when we assume that $c = 2$, we find the same penalty as Mallows [22].

When the weights are all constants, we consider that each model candidate to the estimation have themselves the same complexity. Consequently we assume that the weights are constants, when we have previously selected by some preliminary procedure

the models candidates. As an example, we get constants weights when the models \mathcal{M}_n are constructed by using CART procedure [14]. That preliminary procedure gives us embedded piecewise models (through the maximal tree where each node represents a dyadic segmentation of the model). Because with this strategy all models have the same weight; even if there dimensions are very different.

When the weights are constants, the penalised estimator would become a kind of generalized form of Mallows' penalty. Thus we do not turn the use of concentration inequality to our advantage, let us talk about the variable weights.

Variable weights

An other way to chose the weights consists of varying w_M with the dimension ; $w_M = k_M \cdot w(k_m) > 0$ for a suitable positive function $w(\cdot)$ with respect of the summing of the weights :

$$\Sigma = \sum_{k=1}^{N_n} \#\{M \in \mathcal{M}_n : k_m = k\} \cdot e^{-k \cdot w(k)} < +\infty.$$

Thanks to $\log \binom{N_n}{k} \leq k \cdot \log \frac{e \cdot N_n}{k}$; the following inequalities hold:

$$\begin{aligned} \Sigma &= \sum_{M \in \mathcal{M}_n | k_M > 0} e^{-w_M \cdot k_M} \\ &= \sum_{k=1}^{N_n} |\{M \in \mathcal{M}_n : k_m = k\}| e^{-w(k) \cdot k} \\ &\leq \sum_{k=1}^{N_n} \binom{N_n - 1}{k - 1} e^{-w(k) \cdot k} = \sum_{k=1}^{N_n} \frac{k}{N_n} \binom{N_n}{k} e^{-w(k) \cdot k} \\ &\leq \sum_{k=1}^{N_n} \frac{k}{N_n} \left(\frac{e \cdot N_n}{k}\right)^k e^{-w(k) \cdot k} \leq \sum_{k=1}^{N_n} \left(\frac{e \cdot N_n}{k}\right)^k e^{-w(k) \cdot k} \\ &\leq \sum_{k=1}^{N_n} e^{-k[w(k) - 1 - \ln(N_n/k)]} \end{aligned}$$

When we choose the weights function w namely, $w(k) = 1 + \delta + \ln(N_n/k)$, we would assume that $\Sigma \leq (1 - e^{-\delta})^{-1}$.

Thus the selection theorem gives us the general form of the penalty (2.8), we are able to get the form of the penalty. When the models candidates of \mathcal{M}_n are the set of all disjoint partitions of \mathcal{X} build by using some canonical basis, we propose the following penalty function:

$$\text{pen}(M) = r \cdot k_M \cdot \frac{\sigma^2}{n} \cdot \left[c_1 \log \left(\frac{N_n}{k_M} \right) + c_2 \right] \quad (2.11)$$

where the calibration pair (c_1, c_2) depends on Σ , K and r .

All along the next sections, we will consider that case: the models candidates of \mathcal{M}_n are the set of all distinct partitions of \mathcal{X} build by using elements of M_0 .

2.3.5 Importance of the calibration

Now we have to discuss about the importance of the calibration pair (c_1, c_2) . A fortiori the calibration does not depend on the model. It depends on three elements :

1. The common probability law of the errors $(\varepsilon_i)_{i=1, \dots, n}$;
2. The choice of the weights $(w_M)_{M \in \mathcal{M}_n}$;
3. $K > 1$ some scalar from the selection theorem.

Otherwise the distribution of the errors is given. And we have previously discussed of the choice of the weights (see section 2.3.2). But, K is an unknown scalar. When we look at the selection theorem, we could see that the scalar K comes from the use of several inequalities. We believe that the accumulation of approximation leads to get pessimistic approximation of the general value of K . As it was proposed by the authors of the selection theorem, we will approximate the calibration by using some calibration procedure.

Moreover, the penalty estimation is closely linked with the oracle model. Thanks to the selection theorem, we have an oracle inequality (2.9):

$$\mathbb{E}_f \left[\|f - \tilde{f}\|_n^2 \right] \leq C(c_1, c_2) \cdot \left\{ \inf_{M \in \mathcal{M}_n} \left(\|f - f_M\|_n^2 + \text{pen}_{c_1, c_2}(M) \right) + (1 + \Sigma) \cdot \frac{\sigma^2}{n} \right\},$$

where the term $C(c_1, c_2)$ is defined by the calibration pair (c_1, c_2) . Thus the penalized estimator shall mimic the oracle model. So we choose the value of the calibration pair (c_1, c_2) in order that $C(c_1, c_2)$ keeps close to 1. We are able to link the quality of the penalized least squares estimator with the calibration constants.

There are various strategies of calibration. Here we will talk about two heuristics methods:

- Heuristic of Birgé and Rozenholc [8];
- Heuristic of Birgé and Massart (see [21], section 8.5.2).

On the first hand, Birgé and Rozenholc [8] have proposed a heuristic of calibration in a very practical paper. The principle of the heuristic consists of analyzing the performances of the penalized estimator versus the oracle estimator through a Monte Carlo procedure. At the end of the procedure, we choose the «optimal» calibration pair (c_1, c_2) in order to minimize the risk bound (2.9).

On the other hand, Birgé-Massart (see [21], section 8.5.2) have proposed to study the bias/variance decomposition of the penalized least squares estimator and adapt Mallows' heuristic [22] to the selection theorem. In a few words, we shall choose the calibration constants in such a way that the penalty compensate the empirical risk for models with high dimensions.

2.4 Theoretical calibration procedure

2.4.1 Aim of the procedure

Thanks to the selection theorem, we assume that we look at an exhaustive research of the piecewise Gaussian regression model (defined in [6, 15]). Referring to the previous work, we consider that the following penalty function holds:

$$\text{pen}(M) = r k_M \frac{\sigma^2}{n} \left[c_1 \log \left(\frac{N_n}{k_M} \right) + c_2 \right]$$

where the calibration pair (c_1, c_2) depends on Σ , K and p and they do not depend on the Gaussian case of the model.

Now our purpose is getting the calibration constants c_1 and c_2 that lead to have *nice* penalized estimator. We will proceed in two times. First of all, we will recall the theoretical calibration procedure of Birgé and Rozenholc [8]. After, we will apply this procedure to the piecewise acceleration model. And in a further section, we will discuss about corrections the penalty function in order to have a data-driven selection procedure (see section 2.6).

The theoretical calibration procedure attempts to get a general calibration pair (\hat{c}_1, \hat{c}_2) . Thanks to the selection theorem, we know that the calibration pair exist, but the theorem only gives us a lower bound of the calibration constants (2.9). That is why we shall implement some calibration procedure. Otherwise, the penalty function only depends on the model through the known parameter σ and its dimension. And σ acts on the penalty function as some multiplicative term. So we believe that we can find an optimal value of the calibration pair (c_1, c_2) by using some Monte Carlo procedure for piecewise linear Gaussian models.

2.4.2 Description of the procedure

The goal calibration procedure is studying the risk ratio between the risk of the penalized estimator and an oracle risk adapted for piecewise regression models:

$$F_n(f; c_1, c_2) = \frac{\mathbb{E}(\|f - \tilde{f}_{(c_1, c_2)}\|_n^2)}{\inf_{1 \leq k \leq N_n} \mathbb{E}(\|f - \hat{f}_k\|_n^2)}$$

where $\tilde{f}_{(c_1, c_2)}$ is the penalized estimator associated with c_1 and c_2 and \hat{f}_k is the least-square estimator of the best partition in k pieces. After F_n is estimating by a simple Monte Carlo procedure and we will choose the final calibration \hat{c}_1 and \hat{c}_2 as the ratio F_n is uniformly minimum for all $f \in \mathcal{P}$ and $n \in \mathcal{N}$.

We define the final procedure through the following two steps:

1. For a given n , we would like to get an optimal value (\hat{c}_1, \hat{c}_2) of the calibration constants independently of the model. Let us define \mathcal{P} some set of piecewise function that is characterizing the average of the models. And let \mathcal{C} be a set of

several values of the calibration constants. After let us estimate F_n by calculating its empirical estimation \bar{F}_n for every values of calibration constants $(c_1, c_2) \in \mathcal{C}$ and for all $f \in \mathcal{P}$. Hence we define a new indicator of the quality of the estimation, namely :

$$\bar{M}_n(c_1, c_2) = \sup_{f \in \mathcal{P}} \bar{F}_n(f, c_1, c_2).$$

2. The optimal calibration couple (c_1, c_2) should minimize \bar{M}_n uniformly on n . So defining \mathcal{N} as a set of various sampling sizes, we replicate the step 1 for all $n \in \mathcal{N}$.

At the end of the procedure, we have the set of the values of the indicators:

$$\{\bar{M}_n(c_1, c_2); \quad (c_1, c_2) \in \mathcal{C} \text{ and } n \in \mathcal{N}\}.$$

And referring to the heuristic of calibration, the optimal value of the calibration constant would minimizes $\bar{M}_n(c_1, c_2)$ uniformly over all $n \in \mathcal{N}$.

2.4.3 Calibration for piecewise affine regression

Now let us apply the calibration procedure. For this application section, we have referred to the work of Lebarbier (see [19], chapter 3) who has applied the calibration procedure to piecewise constant Gaussian models. Here we have applying the calibration procedure with piecewise Arrhenius regression models with Gaussian distributions. For M a given partition of the covariable set, the piecewise Arrhenius regression models are characterizing by f such that :

$$f(x) = \sum_{J \in M} (a_{0,J} + a_{1,J}/x) 1_J(x).$$

For the calibration, we have used the following data :

- The sampling set $\mathcal{N} = \{20; 50; 100; 200\}$.
- For a given $n \in \mathcal{N}$, the covariable x is a regular grid of n elements from 300 to 400.
- The largest partition M_0 is building by using elements of $\{x_i : i = 1, \dots, n\} \cup \{+\infty\}$.
- The scale parameter $\sigma = 1$.
- The set \mathcal{P} is composed of 35 simulated piecewise Arrhenius regression models such that $f^{(j)}(x) = \sum_{J \in M^{(j)}} (a_{0,J}^{(j)} + a_{1,J}^{(j)}/x) \cdot 1_{(x \in J)}$ for all $j = 1, \dots, 35$; where
 - $k^{(j)}$ is the realization of some Poisson random variable $\mathcal{P}(6)$.
 - $M^{(j)}$ is a partition with $k^{(j)}$ intervals characterizing by the realization of an uniform law of $(k^{(j)} - 1)$ drawing over the set $\{x_i : i = 1, \dots, n\}$.

$$- a_{0,J}^{(j)} \sim \mathcal{N}(-10, 10^2) \text{ and } a_{1,J}^{(j)} \sim \mathcal{N}(5000, 200^2).$$

- The calibrations variables are :
 - c_1 varies from 0 to 15 and takes 132 values.
 - c_2 varies from 0 to 20 and takes 121 values.
- The number of Monte Carlo samples $N_{sim} = 250$.

The result of the calibration is summarized in the three next figures (2.1), (2.2) and (2.3). In the figure (2.1), we have represented the function $c_1 \rightarrow \bar{M}_n(c_1, c_2)$ at a given value of $c_2 = 1, 2, 3, 3.4, 4$ or 5 and for various $n \in \mathcal{N}$. The figure shows us that:

- The contrast of the curves increases when c_2 increases.
- The shape of every curves seems themselves very close when c_1 is in the locality of $c_1^*(c_2, n)$ (in a first time the curves are close, and in a second time it increases with c_2 is larger).
- The value of the minimum $c_1^*(c_2, n)$ are themselves close when for medium values of c_2 (around 3).

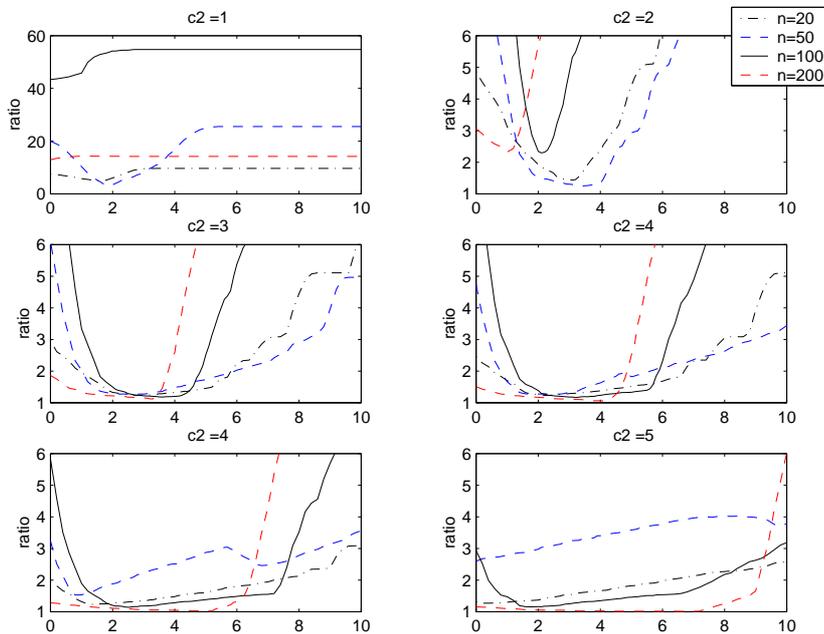


Figure 2.1: The empirical estimation of the risk ratio for piecewise Arrhenius accelerating Gaussian model : we have plot out $\bar{M}_n(c_1, c_2)$ when $c_2 = 1, 2, 3, 3.4, 4$ and 5 in function of c_1 .

n	$\min \bar{M}_n(c_1, c_2)$	$\bar{M}_n(2.8, 3.4)$	$\bar{M}_n(2.1, 2)$	$\bar{M}_n(3.2, 3)$	$\bar{M}_n(1.4, 4)$	$\bar{M}_n(2, 5)$
20	1.2279	1.2651	1.8918	1.2487	1.2318	1.3681
50	1.2319	1.2839	1.4838	1.3042	1.6544	2.9764
100	1.1381	1.1941	2.2956	1.208	1.5273	1.1633
200	1	1.1189	6.0536	1.1447	1.1433	1.0535

Table 2.1: This table gives us the value of $\bar{M}_n(c_1, c_2)$ when n varies and for some pairs (c_1, c_2) candidates to the calibration.

Now let us discuss about the values of $c_1^*(c_2) = \arg \min_{c_1 \geq 0} [\bar{M}_n(c_1, c_2)]$ for a few values of $c_2 = 2, 3, 4$, and let us compare them with Lebarbier's calibration $(2, 5)$. And let us calculate the global minimum upon the maximum the risk ratio $\{\bar{M}_n(c_1, c_2) - \inf_{(c'_1, c'_2) > 0} (M_n(c'_1, c'_2))\}$. This function is minimum when $c_1 = 2.8$ and $c_2 = 3.4$. We find that :

$$c_1^*(2) = 2.1, c_1^*(3) = 3.1 \text{ and } c_1^*(4) = 1.4.$$

The table 2.1 shows us the behaviour (in terms of the risk ratio) of those calibration pair when n varies. We have taken as a reference $\min \bar{M}_n(c_1, c_2)$, the minimum values of the risk ratio for all sampling sizes. We can see that for $(2.8, 3.4)$, the values of the ratio keeps close to $\min \bar{M}_n(c_1, c_2)$ when n varies. Moreover they are themselves very close (1.19, 1.26 and 1.28).

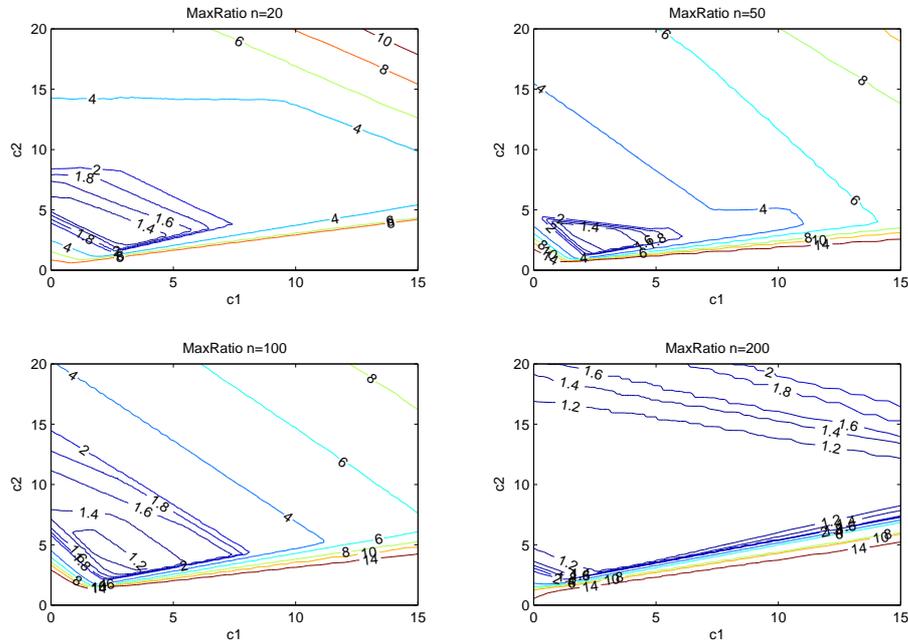


Figure 2.2: Contrast of the risk ratio $\bar{M}_n(c_1, c_2)$ for all $n \in \mathcal{N}$ for Gaussian model.

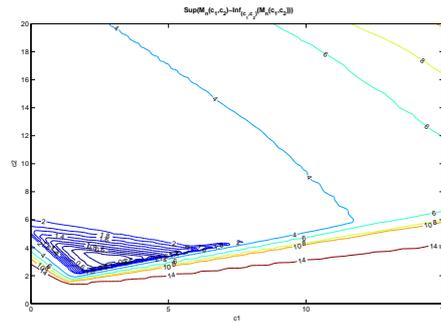


Figure 2.3: Maximum values of contrast of the risk ratio $\{\bar{M}_n(c_1, c_2) - \inf_{(\alpha, \beta) > 0} (M_n(\alpha, \beta))\}$ for Gaussian model.

For the pair (3,3) is quite the same. Notice that looking at figure (2.1), we can hope that the pair (3,3) would have the same kind of behaviour - and it does: $\bar{M}_{20}(3, 3) = 1.2380$, $\bar{M}_{50}(3, 3) = 1.2743$, $\bar{M}_{100}(3, 3) = 1.2193$ and $\bar{M}_{200}(3, 3) = 1.1536$. When $(c_1, c_2) = (1.4, 4)$ or $(2.1, 2)$ or $(2, 5)$, the variability of the risk ratio increases, and some ratio are larger than 2.

So thanks to the table 2.1 and figures (2.1), (2.2) and (2.3), we can assume that the risk ratio is uniformly minimal in a thin area where $c_2 \approx 3.4$ and $c_1 \approx c_2$. In a way of simplification, we propose to use the calibration pair (3,3), so we assume that the penalty function is as follows:

$$\text{pen}(M) \simeq \lambda \cdot r \cdot k_M \cdot \frac{\sigma^2}{n} \cdot \left[\log \left(\frac{N_n}{k_M} \right) + 1 \right], \quad (2.12)$$

where $\lambda \approx 3$. This first calibration will be confirmed by the next experiments.

2.5 Data-driven calibration procedure

In the present section, we will talk about some data-driven calibration based on the heuristic of Birgé-Massart (in details on 2.5.1) and its application to piecewise acceleration models on 2.5.2. The calibration procedure is called data-driven, because we calibrate the penalty function on the data. In a first time, we have described the heuristic. Next we have illustrated (and tested) the calibration procedure for simulated data. And at the end, we have used the heuristic of Birgé and Massart in order to implement a theoretical calibration procedure based on full noisy data.

2.5.1 Birgé-Massart's heuristic

The heuristic of Birgé-Massart (see [21] on section 8.5.2) deals with an update of the heuristic of Mallows [22]. The principle of Mallows' heuristic consists of studying the behaviour of the risk of the least squares estimator \hat{f}_M . Thanks to Cochran's theorem, we decompose the risk of \hat{f}_M as follows:

$$\begin{aligned}\mathbb{E}(\|f - \hat{f}_M\|_n^2) &= \|f - f_M\|_n^2 + \mathbb{E}(\|f_M - \hat{f}_M\|_n^2) \\ &= \|f - f_M\|_n^2 + \sigma^2 \frac{r \cdot k_M}{n}\end{aligned}$$

where f_M is the projection of f to the linear model \mathcal{F}_M with respect to the norm $\|\cdot\|_n$. Looking at the risk, the best partition M^* is the one which is minimizing:

$$\|f - f_M\|_n^2 + \sigma^2 \frac{r \cdot k_M}{n}.$$

As f does not depend on M , M^* minimizes the next criterion:

$$-\|f_M\|_n^2 + \sigma^2 \frac{r \cdot k_M}{n}.$$

The key of Mallows' heuristic consists of replacing the term $\|f_M\|_n^2$ by an unbiased estimator. An *ad-hoc* estimator of $\|f_M\|_n^2$ is $\|\hat{f}_M\|_n^2$ but it is biased:

$$\begin{aligned}\mathbb{E}(\|\hat{f}_M\|_n^2) &= \|f_M\|_n^2 + \mathbb{E}(\|\hat{f}_M - f_M\|_n^2) \\ &= \|f_M\|_n^2 + \sigma^2 \frac{r \cdot k_M}{n}.\end{aligned}$$

So we shall estimate $\|f_M\|_n^2$ by $\|\hat{f}_M\|_n^2 - \sigma^2 \frac{r \cdot k_M}{n}$ an unbiased estimator.

Now consider $k = 1, \dots, N_n = n$, and defining M_k as it minimizes $-\|\hat{f}_M\|_n^2$ when $k_M = k$. Furthermore, let \hat{f}_k be the least squares estimator associated with the partition M_k the best partition on k pieces. Mallows proposes to minimize the following term:

$$-\|\hat{f}_k\|_n^2 + 2\sigma^2 \frac{r \cdot k}{n}. \quad (2.13)$$

The heuristic of Mallows supposes that $\|\hat{f}_k\|_n^2$ keeps close to its expectation for all M simultaneously. But this assumption is only true when the partition M is not too large. However this phenomenon has intrinsically acted to the choice of the penalty function of Birgé and Massart [6]. Thanks to the selection theorem, we take care of the complexity for large models on the selection criterion, and we replace $\sigma^2 \frac{r \cdot k}{n}$ by $\sigma^2 \frac{r \cdot k}{n} [c_1 \log(N_n/k) + c_2]$.

The heuristic of Birgé-Massart supposes that $\|\hat{f}_k\|_n^2$ keeps close to its expectation and half of the penalty term:

$$\|\hat{f}_k\|_n^2 \simeq \mathbb{E}(\|\hat{f}_k\|_n^2) = \|f_k\|_n^2 + \frac{1}{2} pen(M_k).$$

Moreover, the term $\|f_k\|_n^2$ is constant when k is large. Thus $\|f_k\|_n^2 = \|f\|_n^2 + \|f - f_k\|_n^2$ by definition. At the same time, when M is large $\|f - f_k\|_n^2 \simeq 0$. At last, when k is large enough, we assume that :

$$\begin{aligned} -\|\hat{f}_k\|_n^2 &\simeq \frac{1}{2} \text{pen}(M_k) + c_0 \\ &\simeq \frac{1}{2} r k \frac{\sigma^2}{n} c_1 \left[\log\left(\frac{n}{k}\right) + \frac{c_2}{c_1} \right] + c_0 \end{aligned} \quad (2.14)$$

and let us denoting $\lambda = c_1$ and $c = c_2/c_1$.

Now let us use this heuristic in order to calibrate the penalty function. The penalized criterion have to select the \hat{k} number of pieces of the best partition that minimizes the penalized criterion (2.5). Thanks to Birgé-Massart's heuristic, when $k_{min} \leq k \leq k_{max}$, we estimate the calibration pair (c_1, c_2) through a least squares fit of $-\|\hat{f}_k\|_n^2$ on $\frac{1}{2} r k \frac{\sigma^2}{n} \lambda [\log(n/k) + c] + c_0$.

Remark 2.5.1. If σ is unknown, we are able to use the heuristic of Birgé-Massart. We just have to integrate this unknown parameter to the fit, and then we have to realize the following regression:

$$-\|\hat{f}_k\|_n^2 \simeq \frac{1}{2} \frac{r k}{n} \lambda' [\log(N_n/k) + c] + c_0, \quad \forall k = k_{min}, \dots, k_{max}.$$

2.5.2 Application of Birgé-Massart's heuristic

Now, we have estimated the median of the calibration constants through a Monte Carlo procedure using 100 simulated samples of by procedure a 3 Arrhenius Lognormal ALT for various sampling sizes $n = 50, 100, 200$ and 400. The accelerating function is build by using 3 piecewise Arrhenius, as:

$$f(x) = \begin{cases} -24 + 12900/x, & \text{if } x < 345; \\ -145 + 52900/x, & \text{if } 345 \leq x < 370; \\ -46 + 13400/x, & \text{if } 370 \leq x. \end{cases}$$

Let $\sigma = 0.8$ and the covariable vector is a regular vector x , when $n=50$ (resp $n=100$ and so on), \mathbf{x} is a vector defined as a regular grid of the interval $[300;400]$ in 50 (resp 100, and so on). Next we have tested various values of k_{min} and k_{max} , that are two parameters of the heuristic. We have compared the following ranges for the high dimensional models:

- $k_{min} = 5$ or 10;
- $k_{max} = n/4, n/3$ or $n/2$.

The simulated study is summarizing by two tables 2.2 and 2.3, and two figures 2.4 and 2.5. The table 2.2 gives us the MC-estimation of c , λ and the empirical risk ratio between the estimation risk and the oracle risk of grouped models. We could see that the estimation of λ seems stable and does not vary with k_{min} or k_{max} . This

(k_{\min}, k_{\max})	$(5, n/4)$	$(5, n/3)$	$(5, n/2)$	$(10, n/4)$	$(10, n/3)$	$(10, n/2)$
$n=50: \hat{c}$	1.1940	1.1758	1.1454	1.0417	1.1387	1.1164
$\hat{\lambda}$	1.6284	1.6089	1.5918	1.4887	1.5772	1.5788
risk ratio	1.8314	1.8314	1.8314	13.9644	2.6490	1.8314
$n=100: \hat{c}$	1.1890	1.1898	1.1383	1.1215	1.1300	1.1205
$\hat{\lambda}$	0.8194	0.8409	0.8436	0.8325	0.8346	0.8268
risk ratio	1.8721	1.8341	1.9429	2.4854	2.2681	2.5659
$n=200: \hat{c}$	1.1302	1.1540	1.1239	1.1531	1.1455	1.1254
$\hat{\lambda}$	1.2408	1.2380	1.2382	1.2409	1.2389	1.2390
risk ratio	1.0274	1.0274	1.0274	1.0274	1.0274	1.0274
$n=400: \hat{c}$	1.1551	1.1646	1.1414	1.1556	1.1791	1.1352
$\hat{\lambda}$	1.8441	1.8752	1.8929	1.8436	1.8752	1.8924
risk ratio	1	1	1	1	1	1

Table 2.2: The table represents the median and mean estimations of c^* for Lognormal ALT with various sampling sizes.

(k_{\min}, k_{\max})	$(5, \frac{n}{4})$	$(5, \frac{n}{3})$	$(5, \frac{n}{2})$	$(10, \frac{n}{4})$	$(10, \frac{n}{3})$	$(10, \frac{n}{2})$
$n=50: k$	2 3 4 5	2 3 4 5	2 3 4 5	2 3 4 5	2 3 4 5	2 3 4 5
$\bar{p}_{\hat{k}}(k)$	0 100 0 0	0 100 0 0	0 100 0 0	8 58 1 0	0 91 1 0	0 100 0 0
$n=100: k$	2 3 4 5	2 3 4 5	2 3 4 5	2 3 4 5	2 3 4 5	2 3 4 5
$\bar{p}_{\hat{k}}(k)$	0 76 11 2	0 76 13 1	0 74 13 2	0 72 10 1	0 72 12 1	0 68 9 6
$n=200: k$	2 3 4 5	2 3 4 5	2 3 4 5	2 3 4 5	2 3 4 5	2 3 4 5
$\bar{p}_{\hat{k}}(k)$	0 99 1 0	0 99 1 0	0 99 1 0	0 99 1 0	0 99 1 0	0 99 1 0
$n=400: k$	2 3 4 5	2 3 4 5	2 3 4 5	2 3 4 5	2 3 4 5	2 3 4 5
$\bar{p}_{\hat{k}}(k)$	0 100 0 0	0 100 0 0	0 100 0 0	0 100 0 0	0 100 0 0	0 100 0 0

Table 2.3: The table represents $\bar{p}_{\hat{k}}(k) = \frac{1}{100} \sum_{j=1}^{100} 1_{(\hat{k}(y^{(j)}, \mathbf{x}^{(j)})=k)}$ the empirical distribution of k using the PLSE for Lognormal ALT with various sampling sizes.

observation is confirming by the box plot of the estimation of λ on figure 2.5. Such a result is logical, as λ comes from a complexity term. Otherwise, the estimation of c is not constant when n , k_{\min} or k_{\max} are varying. But looking at the box plot on figure 2.4, we could see that for a fixed value of k_{\min} the box plots are nested when k_{\max} varies. Also we could see that the variability of estimations of c is decreasing when k_{\max} is increasing and for a fixed k_{\min} . The choice of n , k_{\min} and k_{\max} have an effect on the estimation of c .

Now let us discuss about the quality of the selection method. The table 2.2 shows that the risk ratio is decreasing when n is increasing and it is close to 1 when $n \geq 200$. And when $n \leq 100$, the risk ratio are not large when $k_{\min} = 5$ and they are decreasing with k_{\max} when $k_{\min} = 10$.

Next, the table 2.3 gives us the empirical distribution of \hat{k} (the estimation of k the number of ruptures). The empirical distribution of \hat{k} gives the same kind of result as we have previously observed. If $n \leq 200$, than the PLSE returns almost surely the right number of ruptures. After, if $n \leq 100$ and $k_{\min} = 5$, the success rate is constant.

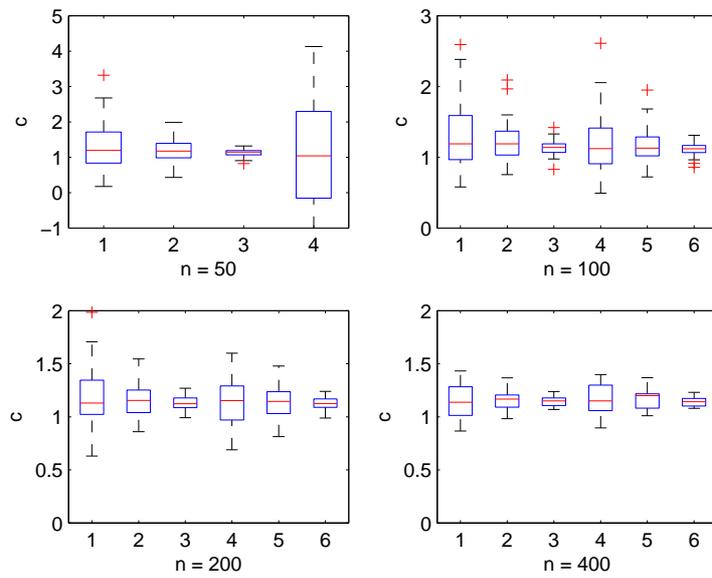


Figure 2.4: Boxplot for c for various series of sampling sizes, in $1 \leq k \leq n/4$, in serie $2 \leq k \leq n/3$, ... and n varies from 50 to 400.

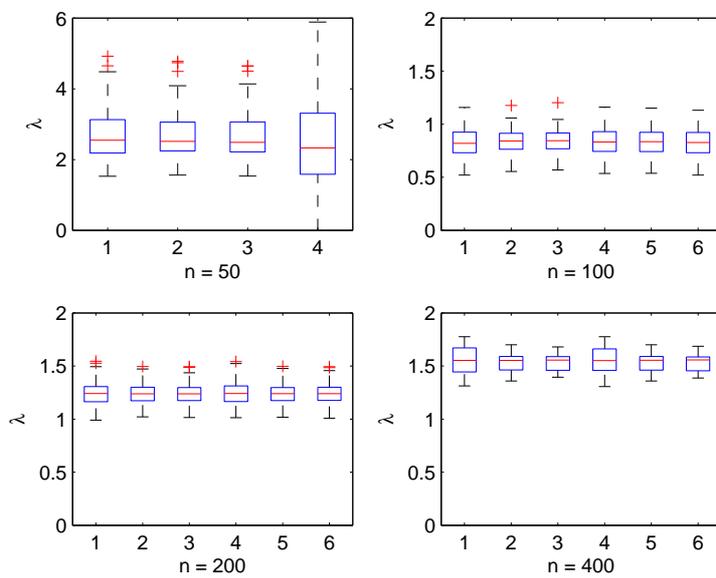


Figure 2.5: Boxplot for λ for various series of sampling sizes, in $1 \leq k \leq n/4$, in serie $2 \leq k \leq n/3$, ... and n varies from 50 to 400.

And if $n \leq 100$ and $k_{\min} = 10$, the success rate is higher when $k_{\max} < n/2$.

At the end, the first method gives us a simple selection method if n is larger than 200. And if n is not large, we should take care of the values of the couple (k_{\min}, k_{\max}) ; especially we have to get k_{\min} which is not too far from the *real* value of k or a large k_{\max} .

2.5.3 Birgé-Massart's heuristic on full noisy data

In this section, we will discuss of the use of Birgé-Massart's heuristic in order to get the calibration constants. The main idea of the following calibration procedure consists of simulating some full noisy data and also applying Birgé-Massart's heuristic (see [21] on section 8.5.2). If $\mu = 0$ the data are full noisy data and then there is no statistical information. The term $\mathbb{E}[\|\hat{f}_k\|_n^2]$ is approximately constant for all piecewise affine models made of k pieces. And then, the heuristics assumes that:

$$-\|\hat{f}_k\|_n^2 \simeq \frac{1}{2} \frac{pk}{n} \lambda [\log(n/k) + c] + c_0, \quad \forall k \leq k_{\max}.$$

So a simple and quick calibration procedure consists of estimating c based on a Monte Carlo method and using Birgé-Massart's heuristic for noisy data. And thanks to this simulating study, the form of the penalty would be:

$$\text{pen}(M) = \lambda^* \sigma^2 \frac{rk}{n} \left[\log \left(\frac{N_n}{k} \right) + c^* \right],$$

where the calibrations constants λ^* and c^* are given by the Monte Carlo procedure.

Procedure

The simple procedure of calibration of c will be a Monte Carlo study based on 100 noisy frameworks for the standard distributions $N(0, 1)$. Let $(Z^{(j)})_{j=1, \dots, 100}$ be these 100 frameworks. For each regression data $Z^{(j)}$, we have to calculate $M_{k^{(j)}}$ and $\hat{f}_{k^{(j)}}$. Then we shall estimate the pair $(\lambda^{(j)}, c^{(j)})$ by the following linear estimation:

$$-\|\hat{f}_k\|_n^2 \simeq \lambda^{(j)} \frac{rk}{n} \left[\log \left(\frac{n}{k} \right) + c^{(j)} \right] + c_0.$$

For Gaussian distributions, we have realized the simulating calibration procedure for various sampling sizes $n = 50, 100, 200$ and 400 , $k_{\min} = 1$ and $k_{\max} = n/5, n/4, n/3$ or $n/2$. And we have considered different covariables vector T for each sampling size. When $n=50$ (resp $n=100, n=200$, and so on), T concatenate one vector defined as some regular grid of the interval $[300;400]$ in 50 (resp 100, 200, and so on) values and k_{\max} varies.

In the table 2.4, we have regrouped the median, the mean and the standard deviation based on the estimations of c when $n=50, 100, 200$ or 400 and $k_{\max}=n/5, n/4, n/3$ or $n/2$. For a given sampling size, we see that the standard deviation are clearly decreasing when k_{\max} is increasing. Otherwise since n is fixed, the median and the mean are near

Table 2.4: The table represents the median and mean estimations of c^* based on the pure noisy method for Lognormal ALT with various sampling sizes.

	K_{\max}	$n/5$	$n/4$	$n/3$	$n/2$
$n=50$:	median	1.1152	1.0907	1.1182	1.1032
	mean	1.3772	1.1977	1.1321	1.0926
	std	1.0028	0.5187	0.2979	0.1319
$n=100$:	median	0.9953	1.0291	1.0925	1.0894
	mean	1.2377	1.1464	1.1181	1.0979
	std	0.7008	0.4207	0.2311	0.0944
$n=200$:	median	1.0700	1.0543	1.0958	1.0963
	mean	1.0989	1.0892	1.1057	1.1058
	std	0.3542	0.2470	0.1582	0.0725
$n=400$:	median	0.9385	1.0139	1.0939	1.1106
	mean	0.9614	1.0140	1.0772	1.1081
	std	0.2127	0.1549	0.1063	0.0479

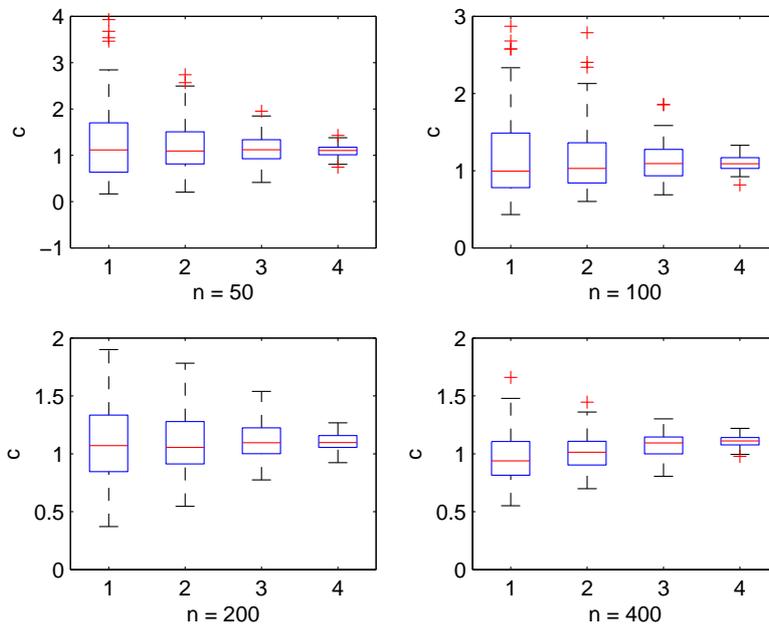


Figure 2.6: Boxplot for c for various series of sampling sizes, in 1 $n = 50$, in 2 $n = 100$, in 3 $n = 200$ and in 4 $n = 400$ and k_{\max} varies from $n/5$ to $n/2$.

the same value when $k_{\max}=n/2$; so the empirical distributions of \hat{c} seems symmetric. Next, looking at the box plot of c on figure 2.6, we observe that the box plots of c are nested for n fixed. The figure also show us that the variability of the estimations decreases when k_{\max} increases.

Finally, when $k_{\max} = \frac{n}{2}$, the distribution of the Monte-Carlo estimators are clearly concentrating around the median (and the mean), and for all n , the estimator of c^* are round the same value. So we propose to get $c^* = 1.1016$, which is te average of the median of c when $n=50, 100, 200$ and 400 .

2.6 Data-driven correction of the calibration

During this section, we will consider that the form of penalty function is given as follows:

$$\begin{aligned} pen(M) &= \alpha g_n(k) \\ &= \alpha r k_M \frac{\sigma^2}{n} \left[c_1 \log \left(\frac{n}{k_M} \right) + c_2 \right]. \end{aligned} \quad (2.15)$$

where the calibration pair is given by the procedure of Birgé and Rozenholc (see section 2.4) or using the heuristic of Birgé and Massart on full noisy data where $c_1 = \lambda$ and $c_2 = c.\lambda$ (see section 2.5.3).

2.6.1 Aims and method of correction

In practice, we usually correct the theoretical calibration constants, thus working with real data leads to additional constraints :

- Misleading us on the scale of the penalty function ;
- Unknown scale parameter σ .

First of all, it is clear that some misleading of the scale of the penalty function leads to overestimate the complexity of the *real* model (as it does for AIC). Moreover, the theoretical calibration procedures are Monte Carlo procedures, some people would like to adapt the penalty term to the data. When the calibration constants are misrepresenting the data ; we have to find where the heuristics of calibration mismatches the data. That is why we will use once again the heuristics of calibration : in order to correct the calibration pair by using some multiplicative scalar to the penalty. Furthermore, the correction procedure of the calibration gives us a practical data-driven method.

On the other hand, the scale parameter σ maybe unknown. Unfortunately, we have supposed that σ is known when we have constructed the penalized criterion (referring to assumption of the selection theorem). However, the scale parameter σ only influences the penalty function through some multiplicative term σ^2 . Thanks to that statement, we could integrate the unknown σ into the correction term.

Remark 2.6.1. Let us notice that Huet have worked on variable selection with Gaussian and when σ is unknown. She has used Kullback-Liebler distance instead of quadratic risk and she has proposed a selection theorem using the same kind of penalty function as Birgé and Massart do.

Anyway, we generally consider that the penalty function is multiplying by some corrective term α . Thanks to the general study of Birgé and Massart [6, 7, 21], we assume that when the value of α goes to infinity, the bound of the oracle inequality (2.9) holds, but the inequality also becomes larger for larger values of α . Next, we assume that when the value of α is less than 1, the selected model is overestimating the dimension of the *true* model. In this work, we have referred to the thesis of Lebarbier [19], who has applied the heuristics of Birgé and Massart in order to detect the ruptures on some piecewise constant Gaussian data. When σ is unknown, we estimate it in order to get an idea of its magnitude ; as Lebarbier we have use the non-parametric estimator of Hall and al. (1990). Then, we have to find the minimal calibration term α_{\min} using an heuristic of calibration. And we estimate the final correction term $\alpha_{\text{opt}} = 2\alpha_{\min}$. At the end, we are able to estimate the PLSE based on data-driven selection method.

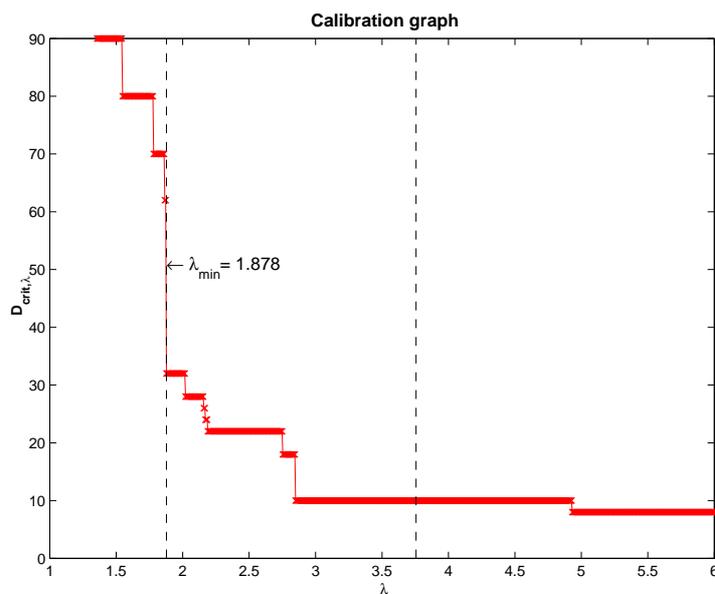


Figure 2.7: Representing the variations of the penalized estimation when the correction term α varies.

For some fixed value of $0 \leq \alpha$, let \hat{k}_α be the number of ruptures detecting. It is easy to prove (see [19], lemma 4.4.1 and 4.4.2.) that there is a sequence of scales $\alpha_0 = 0 < \alpha_1 < \dots < \alpha_N$ and there is a sequence of numbers $k_0 = N_n > k_1 > \dots > k_N$ such that for all $\alpha \in [\alpha_j; \alpha_{j+1}[$ the estimation leads to select $\hat{k}_\alpha = k_j$. The function of the number of ruptures $\alpha \rightarrow \hat{k}_\alpha$ is decreasing. It is step function based on $(k_j)_{1, \dots, N}$. And $-\alpha_j$ is the coefficient estimating when we get the regression of $\|\hat{f}_k\|_n^2$ by $g_n(k)$. In practice, when we plot the step function $\alpha \rightarrow \hat{k}_\alpha$, we would observe some

significant rupture on dimension of the model that underlines the value of α_{\min} the minimal penalty. As an example, we have represented this step function on the figure 2.7 for some Lognormal accelerating life test based on an acceleration function build with 4 Arrhenius acceleration steps.

In the example given on figure 2.7, we clearly see some significant dimensional gap on g_n . The minimal penalty is defined as the value of the largest dimensional gap, denoting α_{\min} . At last we define the optimal penalty denoting α_{opt} as $\alpha_{opt} = 2.\alpha_{\min}$. When the form of the penalty was correctly calibrated, the value of the largest dimensional gap shows us the α_{\min} . And if the largest models leads to poor or even non convergent estimation, than the value of the largest gap may induce a small α_{\min} . In such case, we notice that the second larger gap is associated with a larger α_{\min} . We could easily see these phenomena in practice, but most of usefully statistical method is automatic. In order to avoid the limits of the method, Lebarbier have proposed to constraint the α_{\min} to be larger that $\sigma^2 \beta_{\text{seuil}}$. In practice, σ is known or estimated and the scalar β_{seuil} is estimating though a test which detects if there is not any rupture :

$$H_0 : (\alpha \geq \beta_{\text{seuil}} \hat{\sigma}^2) \text{ versus } H_1 : (\alpha < \beta_{\text{seuil}} \hat{\sigma}^2).$$

Thanks to a study of the empirical distributions of α with Gaussian frameworks, Lebarbier have proposed to get $\beta_{\text{seuil}} = 0.62$ when $n < 200$, and $\beta_{\text{seuil}} = 0.76$ when $n \geq 200$.

2.6.2 Numerical application

We have performed the penalized estimator by using the data-driven correction of the penalty with the same example as we have told in section 2.5.2 :

$$f(x) = \begin{cases} -24 + 12900/x, & \text{if } x < 345; \\ -145 + 52900/x, & \text{if } 345 \leq x < 370; \\ -46 + 13400/x, & \text{if } 370 \leq x. \end{cases}$$

During the selection procedure, we have supposed that σ is known. As [19] we have used the non parametric estimator of Hall and al. in order to get an approximation of σ . However, we can also choose the largest estimator of σ over all plausible piecewise models. Next, we have applied the method using the calibration procedure of Birgé and Rozenholc (denoting PLSE1) and we have also applied the method with the procedure based on full noisy data (denoting PLSE2). Furthermore, we have implemented the classical penalized estimators AIC, BIC and AICc. And lastly, we have tested the PLSE in order to correct the penalty function of BIC with the previous data-driven procedure where $g_n(k) = k \log(n)$. We have denoted BICc that data-driven criteria based on the penalty of BIC.

	PLSE1	PLSE2	AIC	AICc	BIC	BICc
$n = 50 : k$	2 3 4 5	2 3 4 5	2 3 4 5	2 3 4 5	2 3 4 5	2 3 4 5
$\bar{p}_{\hat{k}}(k)$	0 71 9 6	0 65 11 4	0 5 9 22	0 0 0 0	0 61 27 7	0 77 16 6
$n = 100 : k$	2 3 4 5	2 3 4 5	2 3 4 5	2 3 4 5	2 3 4 5	2 3 4 5
$\bar{p}_{\hat{k}}(k)$	0 81 13 2	0 75 14 3	0 0 4 3	0 0 0 0	0 68 25 4	0 73 18 5
$n = 200 : k$	2 3 4 5	2 3 4 5	2 3 4 5	2 3 4 5	2 3 4 5	2 3 4 5
$\bar{p}_{\hat{k}}(k)$	0 91 5 3	0 89 6 4	0 0 0 0	0 0 0 0	0 65 18 12	0 93 6 1
$n = 400 : k$	2 3 4 5	2 3 4 5	2 3 4 5	2 3 4 5	2 3 4 5	2 3 4 5
$\bar{p}_{\hat{k}}(k)$	0 98 2 0	0 96 4 0	0 0 0 0	0 0 0 0	0 68 22 9	0 100 0 0

Table 2.5: Table of the empirical distribution of \hat{k} the penalized estimator of the number of ruptures on the data and for various selection procedures: AIC, BIC, AICc and the adaptation penalty with the three data-driven methods : PLSE1, PLSE2 and BICc.

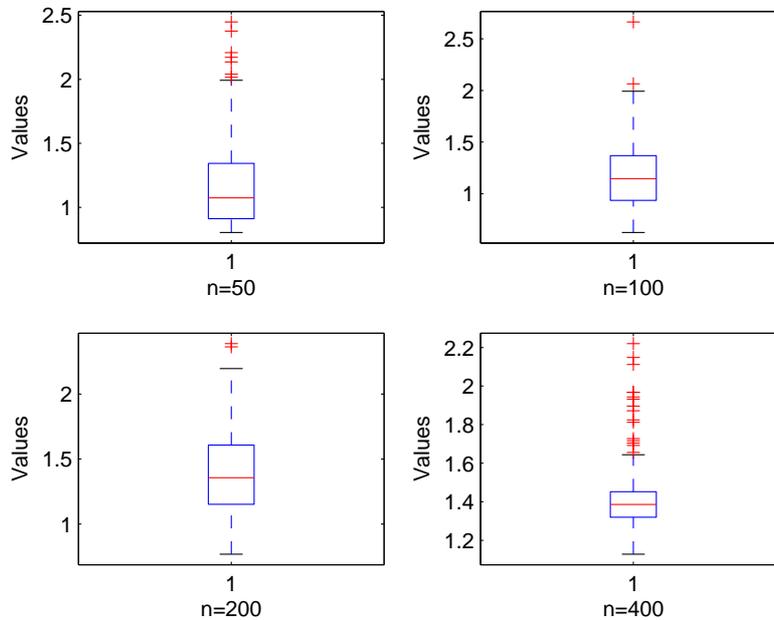


Figure 2.8: Box plot of correction since the penalty function of BIC for all sampling sizes.

On the table 2.5, we have got the empirical estimation of \hat{k} the estimation of the number of accelerating step for all selection methods. The PLSE1 and PLSE2 have the same kind of results, the empirical distribution of k is concentrating on the true value of $k=3$, it never underestimates the model and the rate of overestimation is decreasing with n . The methods based on Akaike's method are strongly underestimating the penalty, and they are always overestimating the model. And the BIC criterion gives better results than AIC, but when n is increasing the recognizing rate of BIC stay between 61 and 68%. Moreover, BICc has the same kind of performances as PLSE1

ratio	PLSE1	PLSE2	AIC	AICc	BIC	BICc
$n = 50$	1.9737	2.2031	4.7593	8.7587	2.0508	1.4814
$n = 100$	1.6662	2.0425	8.1829	16.7133	1.8165	1.5533
$n = 200$	1.5577	1.5167	15.5709	31.9942	2.1252	1.1744
$n = 400$	1.0551	1.1052	44.3593	93.4644	2.9463	1.0000

Table 2.6: Table of the empirical risk ratio between the estimation risk and the oracle risk for all tested procedures: AIC, BIC, AICc and our penalty with the three data-driven methods : PLSE1, PLSE2 and BICc.

and PLSE2. We also notice that the median of the correction term of BICc is increasing with n . The box plots of the correction term on figure 2.8 show that the correction term is concentrating around 1 (classical estimation of BIC's penalty). During the simulations, some correction of BIC are larger than the expectation (and not larger than 2). When $n=50, 100$ the performances of BIC and BICc are quite the same. And, when $n=200$ or 400, BICc has a better recognizing rate (around 80%) than BIC (around 65%).

And on the table 2.6, we have given an estimation of the risk ratio between the PLSE and the oracle of the grouped models (dimension per dimension). As in theoretical results, the selection method based on Sauv e's theorem gives the better results in terms of risk ratio. The risk ratio of PLSE methods is decreasing when n is increasing. And it is closer to 1 when n is increasing. Let us notice that the calibration based on the calibration of Birg e and Rozenholc gives better results than the ones based on full noisy data. Also, the risk ratio of the corrected criteria based on penalty of BIC has the same magnitude as the PLSE based on Birg e and Rozenholc's procedure. And next, the risk ratio for criterion of AIC and AICc are clearly larger than two, even when n is large. And BIC's risk ratio are between 1.8 and 2.8 (the recognizing rate of BIC keeps close to 65% although n is increasing). Lastly, the fact that the estimation based on AIC or BIC are far from the oracle estimation clearly explains why the PLSE of the table 2.5 do not detect the true model since the penalty function is AIC, BIC or AICc.

Based on this application, we see that the PLSE methods give better results in terms of recognizing rate (and in terms of risk ratio). The performances of PLSE methods are themselves very close (here PLSE1 have better recognizing rate and BICc have lower risk ratio). It is interesting to notice that BIC's correction term is concentrating around 1 (the empirical estimation of BIC's penalty), but we also calculate correction that are larger than 1. We also notice, when n is larger than 200, that the median of the BIC's correction term is slowly increasing.

2.6.3 Application on real data

In the last section, we propose to apply the selection procedure on real data. We have tested the method on two different cases. The first example deals with a reliability study of class-B insulation motors. We have accelerating data with a different behaviour between lower and higher stress levels. In facts, the experiment reveals competing fails for higher stress conditions. The second example deals with a reliability study

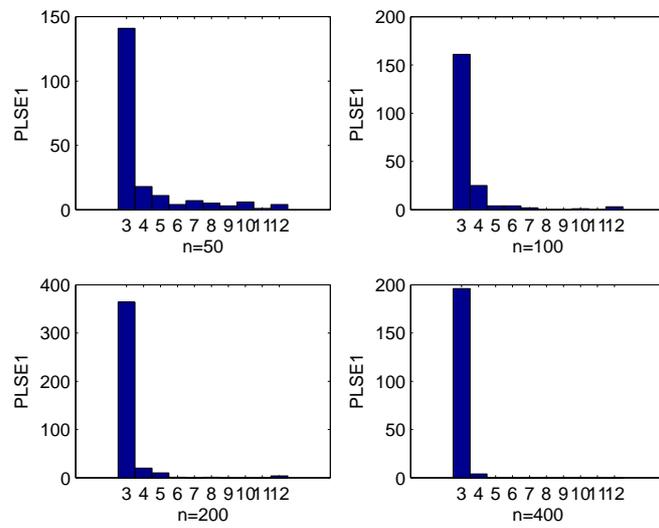


Figure 2.9: The histogramm of \hat{k} based on PLSE1 and for a 3 Arrhenius Lognormal ALT.

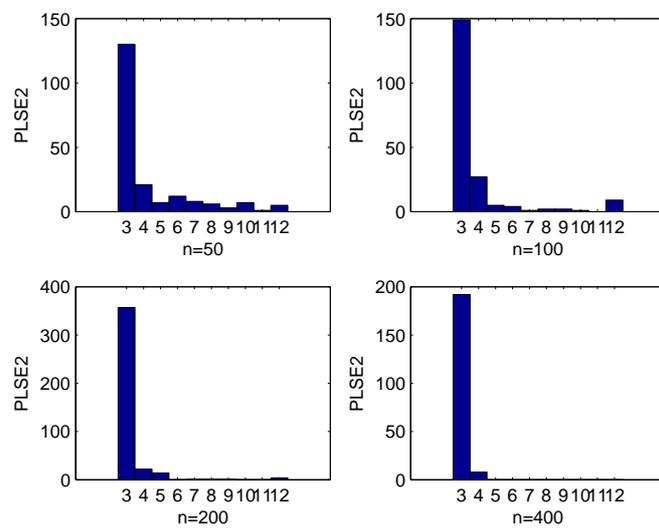


Figure 2.10: The histogramm of \hat{k} based on PLSE2 and for a 3 Arrhenius Lognormal ALT.

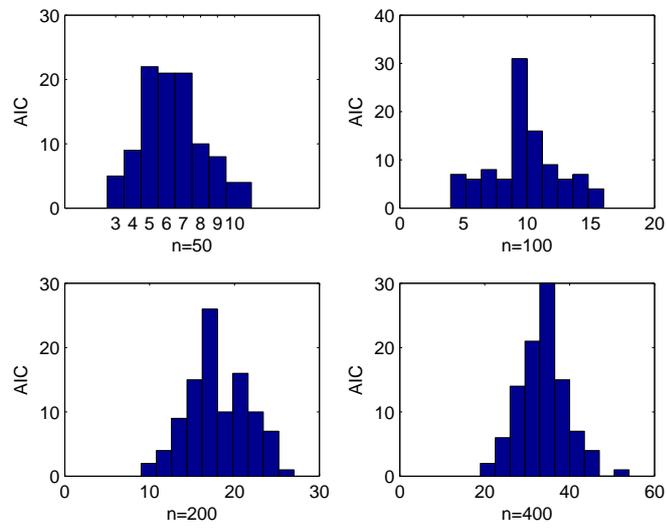


Figure 2.11: The histogramm of \hat{k} based on Akaike's selection method and for a 3 Arrhenius Lognormal ALT.

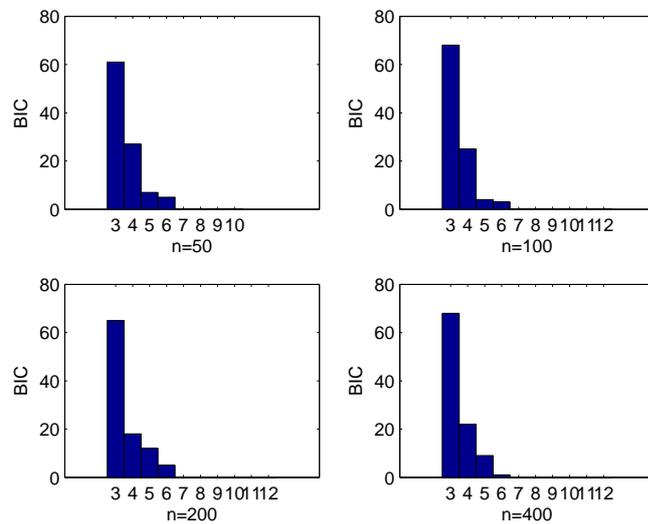


Figure 2.12: The histogramm of \hat{k} based on BIC criteria and for a 3 Arrhenius Lognormal ALT.

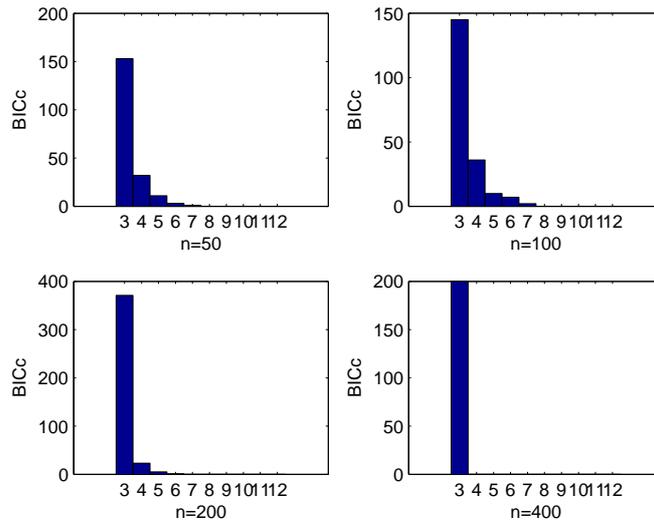


Figure 2.13: The histogram of \hat{k} based on BICc criteria and for a 3 Arrhenius Log-normal ALT.

of battery voltage drop. We have data from a piecewise cubic regression model. The stress sets are given for the experiment, we have applied our model in order to test the detection of ruptures on the cubic regression model.

Example 1 : Class-B motor insulation data

We consider the motor insulation data given by Nelson [27], chapter 7, which is collected from an accelerated testing of motor insulation. The stress used during ALT is temperature (with four temperatures 150, 170, 190 and 220°C) and the acceleration function is Arrhenius. The distribution of the lifetimes is assumed to be Lognormal. The experiment proposes to keep on the test while the first fail occurs. During the test, there is three failure root causes : turn, phase and ground. Nelson proposes to analyse each failure mode independently and to fit a series-system model. Looking at the data, we notice that the ground fail occurs on lower stress levels. Hence the three failure modes are competing on higher stress levels. Here we propose to use a piecewise Arrhenius model with an unknown number of acceleration sets. We forget the root causes and we hope that the selection procedure will detect the two compartments. Such an analysis may not replace a competing risk analysis, our aim is only to underline the misleading on a classical ALT.

The sampling size is 34 (in order to have complete data, we have occulted censored observations that occur at 150°C) because with censored data we should have used MLE instead of LSE. We also assume that $c = 1$, thanks to the estimation of section 2.4.3 based on Birgé-Rozenholc calibration procedure. And we have also consider that the plausible ruptures on the acceleration may occur on temperatures 160, 180 or 200°C. We have run on the dynamic algorithm (see appendix A) and the heuristic

trend (see 2.6.1). The dynamic algorithm gives the following stress sets :

$$\begin{aligned}\hat{M}_1 &= \{ [0, +\infty[\} \\ \hat{M}_2 &= \{ [0, 180[, [180, +\infty[\} \\ \hat{M}_3 &= \{ [0, 160[, [160, 180[, [180, \infty[\} \\ \hat{M}_4 &= \{ [0, 160[, [160, 180[, [180, 200[, [200, \infty[\}\end{aligned}$$

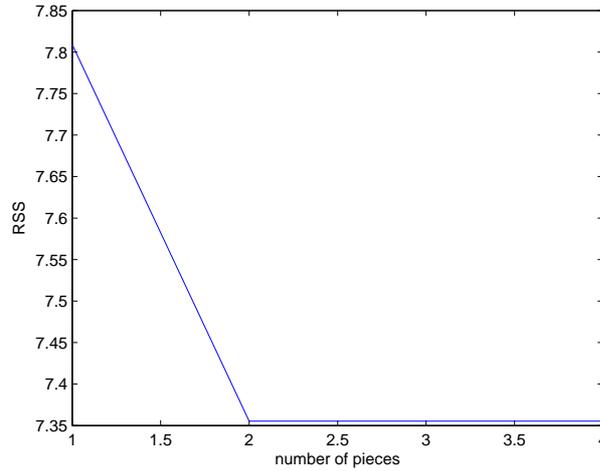


Figure 2.14: The RSS of the piecewise regression models \hat{M}_1 , \hat{M}_2 , \hat{M}_3 and \hat{M}_4 .

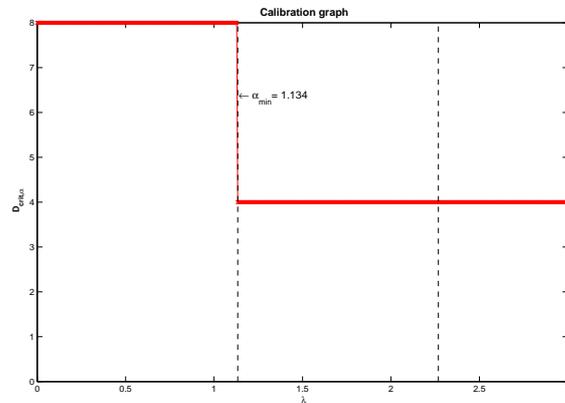


Figure 2.15: The calibration plot represents the selecting dimension when the correcting term α varies.

In figure 2.14, we have traced the residual sum of square of the models candidates to the estimation (indexing by the number of ruptures on the acceleration relationship). Next, the figure 2.14 represents the selecting dimensions versus α the correction term used during the heuristic trend. We observe that the minimal dimensionnal gap occurs when $\alpha_{\min} = 1.134$, so $\alpha_{\text{opt}} = 2.268$ and the selected model is \hat{M}_2 . Such a choice seems

logical, because the selection criterion regroups the stress 150 and 170°C where the ground fail occurs. Otherwise, it regroups the stress 190 and 220°C where we assert competition between the three failure modes. In this example, the selection method have detected the changes between lower an higher stress levels. At the end, we have traced the Class-B data and the PLSE estimation on figure 2.16. And we have also traced the residual plot on figure 2.17.

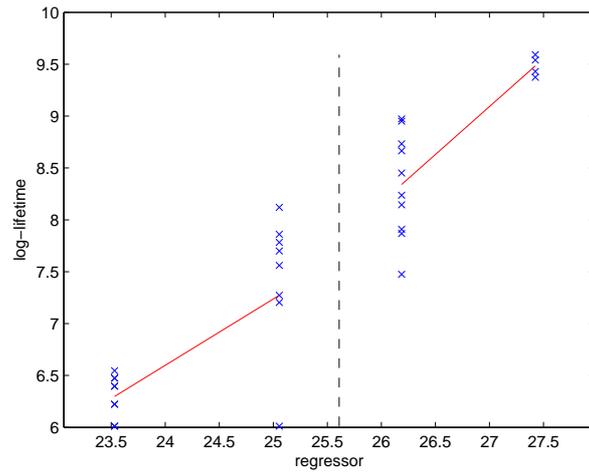


Figure 2.16: The blue crosses represent the observations of class-B insulation data and the red line shows the PLSE.

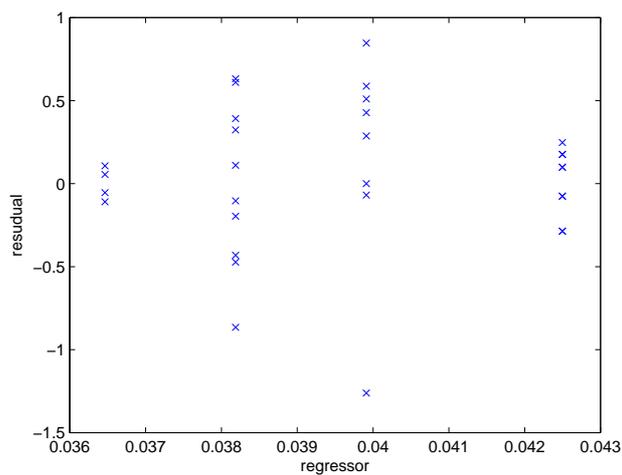


Figure 2.17: The residual plot for PLSE.

Example 2 : Battery voltage drop

Now we consider the battery voltage drop in a guided missile motor observed over the time of missile flight, we have collected these data from the book of Crowder and al. [13] (chapter 5, example 5.2). The regression framework is (y_i, x_i) , where y_i 's are times (in seconds) and x_i 's are voltage drop of the missile's battery. The authors have shown that the course of the missile have changed during the experiment at 6.5 and at 13 seconds. They proposed to fit the data with a cubic spline model, as follows :

$$y = a_{0,0} + a_{0,1}.x + a_{0,2}.x^2 + a_{0,3}.x^3 + a_1.(x - 6.5)_+^3 + a_2.(x - 13)_+^3.$$

Here, we will test our selection procedure in order to detect the number of change on the course of the missile. We propose to fit the data with a piecewise cubic polynomial model, namely :

$$y = \sum_{j=1}^k [a_{j,0} + a_{j,1}.x + a_{j,2}.x^2 + a_{j,3}.x^3] \cdot 1_{(x \in J_k)}$$

where k is the unknown number of ruptures during the flight and (J_k) are distinct time interval which define the time of each course. Here we have consider that k the number or ruptures in less than 10. And we have consider that all ruptures candidates to the estimation are all number from 2 to 18 per 0.5 (noticing that the real ruptures are considered). We have run on the dynamic algorithm (see appendix A) and the heuristic trend (see 2.6.1). The dynamic algorithm gives the following stress sets :

$$\begin{aligned} \hat{M}_1 &= \{ [0, +\infty[\} \\ \hat{M}_2 &= \{ [0, 6[, [6, +\infty[\} \\ \hat{M}_3 &= \{ [0, 6.5[, [6.5, 16.5[, [16.5, \infty[\} \\ \hat{M}_4 &= \{ [0, 5[, [5, 7.5[, [7.5, 16[, [16, \infty[\} \\ \hat{M}_5 &= \{ [0, 5[, [5, 7.5[, [7.5, 14.5[, [14.5, 16.5[, [16.5, \infty[\} \\ \hat{M}_6 &= \{ [0, 5[, [5, 7.5[, [7.5, 14.5[, [14.5, 16.5[, [16.5, 17.5[, [17.5, \infty[\} \\ \hat{M}_7 &= \{ [0, 5[, [5, 6[, [6, 7.5[, [7.5, 14.5[, [14.5, 16.5[, [16.5, 17.5[, [17.5, \infty[\} \\ \hat{M}_8 &= \{ [0, 5[, [5, 6[, [6, 7.5[, [7.5, 8.5[, [8.5, 14.5[, [14.5, 16.5[, [16.5, 17.5[, [17.5, \infty[\} \\ \hat{M}_9 &= \{ [0, 3.5[, [3.5, 5[, [5, 6[, [6, 7.5[, [7.5, 8.5[, [8.5, 14.5[, [14.5, 16.5[, [16.5, 17.5[, \\ &\quad [17.5, \infty[\} \\ \hat{M}_{10} &= \{ [0, 3.5[, [3.5, 5[, [5, 6[, [6, 7.5[, [7.5, 8.5[, [8.5, 13.5[, [13.5, 14.5[, [14.5, 16.5[, \\ &\quad [16.5, 17.5[, [17.5, \infty[\} \end{aligned}$$

The sampling size is 41, we have implemented the method 1 in order to estimate $c = c_2/c_1$, supposing that $K_{\min} = 3$ and $K_{\max} = 10$. In figure 2.18, we have traced the fit of the penalty function. So we find that $c \simeq 1.1718$.

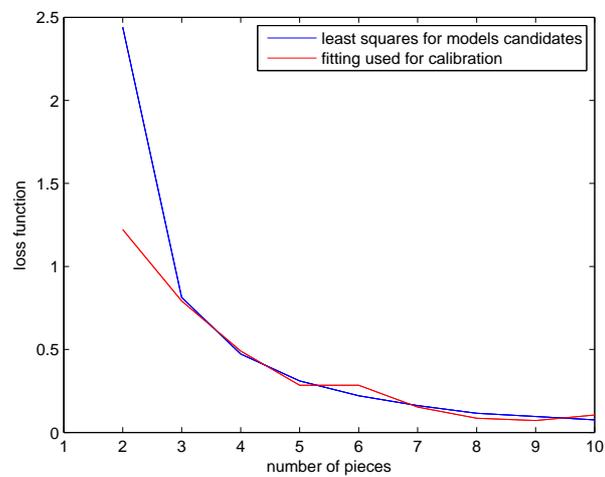


Figure 2.18: The fit of the penalty function as minus half of the euclidean norm of the LSE for $\hat{M}_3, \hat{M}_4 \dots \hat{M}_{10}$

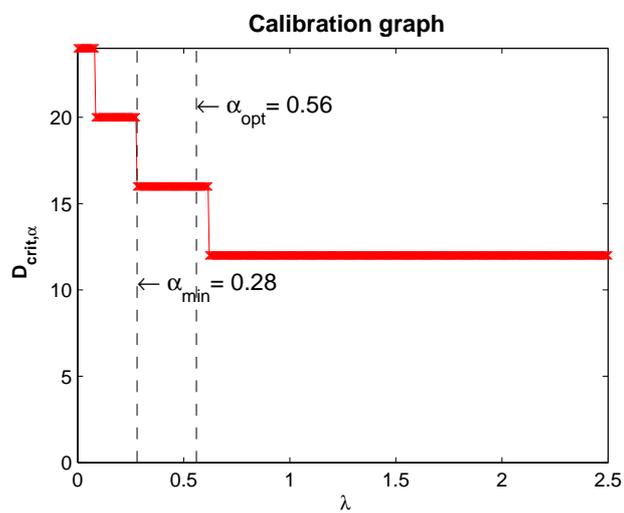


Figure 2.19: The calibration plot represents the selecting dimension when the correcting term α varies.

Next, the figure 2.19 represents the selecting dimensions versus α the correction term using during the heuristic trend. We observe that the minimal dimensionnal gap occurs when $\alpha_{\min} = 0.28$, so $\alpha_{\text{opt}} = 0.56$ and the selected model is \hat{M}_2 . So we find that there is two changes on the course during the flight. Let us notice that the first rupture is correctly detected, but the second one is not. We estimate that the second rupture is 16.5 instead of 13. However the RSS of \hat{M}_2 is close to RSS of the piecewise cubic estimation with the real course (2.1162 instead of 2.2318). The loss of information does not seem determinant for the fit. Hence we show the PLSE fit on figure 2.20 and the LSE based on the real model on figure 2.21. We remark that 13 is close to the second inflection point while 16.5 is placed between the second and the last inflection point of the signal ; it may explain why the second rupture is overestimating. At the end, we have also traced the residual plot for the PLSE on figure 2.22.

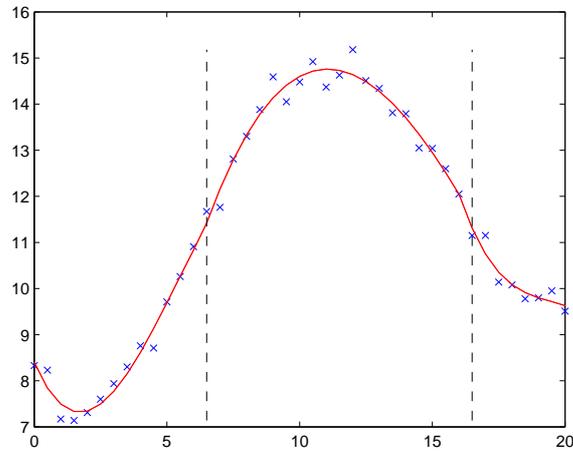


Figure 2.20: The blue crosses represent the observations of voltage drop data and the red line shows the PLSE.

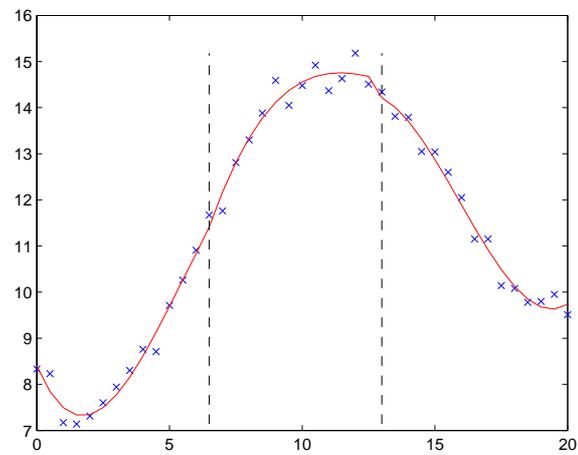


Figure 2.21: The blue crosses represent the observations of voltage drop data and the red line shows the piecewise cubic estimation with two ruptures on the times 6.5 and 13.

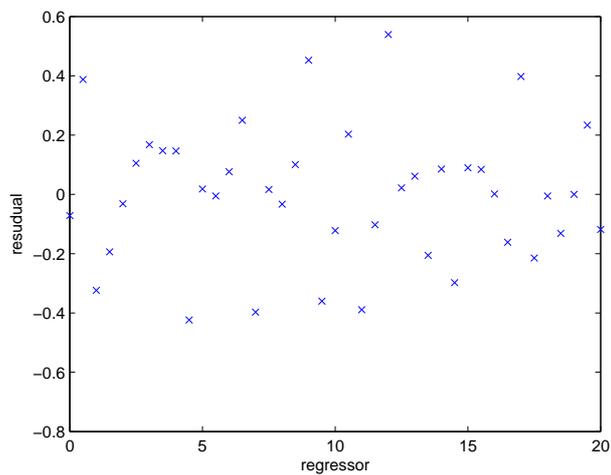


Figure 2.22: The residual plot for PLSE.

Chapter 3

Model Selection for Piecewise Regression Models and Detection of Overstress on Weibull Data

Ce travail a été réalisé en collaboration avec Marie Sauvé.

Abstract : In previous works, we have used Gaussian selection model of Birgé and Massart as a theoretical field and we have estimated the number of acceleration sets of an overstress accelerated test. However, these works do not cover Weibull case which is associating with shock failure. In the present document, we have adapted the recent non-Gaussian selection results of Sauve in order to extend our application field to Weibull case.

Keywords : Acceleration Life Testing, AFT model, Piecewise Regression, Weibull Model Selection, Penalized Criteria, Overstress Accelerating Data.

Résumé : Dans nos précédents travaux, nous avons adapté le cadre théorique de sélection de modèle Gaussien de Birgé et Massart afin d'estimer le nombre de plages d'accélération d'un essai accéléré en sur-stress. Cependant, ces travaux ne recouvrent pas le cas Weibull associé aux défaillances de chocs. Nous avons donc adapté le cadre non-Gaussien, récemment développé par Sauvé, afin d'étendre le champ d'application de notre méthode dans le cas Weibull.

Mots-Clés : Test Accéléré, Modèle AFT, Régression par morceaux, Sélection de Modèle de Weibull, Critère Pénalisé, Données Accélérées en Sur-Stress.

3.1 Introduction

Acceleration Life Tests (ALT) [23, 27] are widely used in reliability fields, because they timely provide lifetime data, even when the tests units are long term components. The principle of ALT consists of testing the lifetime of the test component under higher than usual environmental conditions. The environmental covariables are also called stress. We choose the stress covariables in order to accelerate the observations of the

failure without changing the failure mode. At the end of the experiment, we have timely observed the lifetimes. Then we have to extrapolate the reliability yields through the comportment of the lifetime under stressed conditions. Moreover, we have a better knowledge of the lifetime since the reliability model is enriched with the comportment of the lifetimes when the environmental conditions are varying.

In terms of assumptions for ALT [23], firstly we suppose that the lifetime distribution does not changes when the stress covariable changes. And secondly, there is some relationship which links the stress and the parameters of the lifetime distributions. The kind of life-stress relationship is known but it depends on the parameters which have to be estimated.

As we do in many reliability studies, use for ALT Lognormal or Weibull distribution. And the life-stress relationship links some specific percentile of the lifetime with the stress covariables using the physical, mechanical, chemical or joint effects, which are characterizing the kinetic of the degradation mechanism. We could usually transform the degradation equation as it becomes a linear form. Such a model is called Accelerated Failure Time model, also called AFT models, see the books of Nelson [27] or the recent book of Nikulin and al. [28]. Denoting t_p as the p percentile of the lifetime of the test unit, then the life-stress relationship is some log linear relationship of the transformation of stress covariables, namely:

$$\log(t_p) = a_0 + \sum_{j=1}^{r-1} a_j x^j$$

where x^j is a known transformation of the stress covariables which comes from the linearization of the degradation equation. The life-stress relationship is also called the acceleration function. Many classical life-stress relationships are log linear relationships, and in the following table 3.1 we have listed some classical ones.

Name	Acceleration	Covariables
Arrhenius	$f(x) = a_0 + a_1 \cdot \frac{1}{x}$	\mathbf{x} : temperature
Inverse Power	$f(x) = a_0 + a_1 \cdot \log\left(\frac{1}{x}\right)$	\mathbf{x} : voltage
Eyring	$f(x) = a_0 + a_1 \cdot \frac{1}{x} + a_2 \log\left(\frac{1}{x}\right)$	\mathbf{x} : temperature
Generalized Eyring	$f(x) = a_0 + a_1 \cdot \frac{1}{x^1} + a_2 \log\left(\frac{1}{x^1}\right) + a_3 \cdot x^2 + a_4 \cdot \frac{x^2}{x^1}$	$\mathbf{x} = (x^1, x^2)$: x^1 temperature, x^2 voltage

Table 3.1: Presents a list of non exhaustive life-stress relationship.

Moreover, the choice of the percentile t_p depends on the lifetime distribution. As an example recalling that the Weibull distribution is fully defined by the following probability density function :

$$\psi(t) = \frac{\beta}{\eta} \left(\frac{t}{\eta}\right)^{\beta-1} \exp \left[- \left(\frac{t}{\eta}\right)^\beta \right] 1_{t \geq 0}$$

where $\beta > 0$ is a shape parameter of the model and η is a scale parameter of the model. Recalling that $\eta = t_{63.2\%}$; the acceleration function is supposed to rely the lifetime model and the stress covariables as follows :

$$\log(t_{63.2\%}) = \log(\eta) = a_0 + \sum_{j=1}^{r-1} a_j x^j$$

As β defines the general shape of the lifetimes, referring to the assumptions of ALT, we assume that β is constant and it is independent from the stress covariables. Here we are working on the logarithm of the lifetime, let us recalling that when the lifetime distribution is Lognormal, the log lifetime is Normal; and the logarithm of a Weibull random variable gets a Gumbel random variable with location parameter $\log(\eta)$ and a scale parameter $\frac{1}{\beta}$. If we define $\mu = \log(\eta) - \frac{\gamma}{\beta}$ where $\gamma=0.5722 \dots$ is the gamma Euler's constant and $\sigma = \frac{\pi}{\sqrt{6}} \frac{1}{\beta}$; so let Y be the random variable of the log lifetime, the acceleration model can be written as the following regression problem :

$$Y = \mu + \sigma \varepsilon \quad (3.1)$$

and

$$\mu = a_0 + \sum_{j=1}^{r-1} a_j x^j \quad (3.2)$$

where ε is a random variable of the errors (centred and standardized); when the distribution of the lifetime is Weibull, ε is a centred and standardized transformation of a Gumbel random variable, as $\frac{\pi}{\sqrt{6}} \varepsilon - \gamma \sim SEV(0, 1)$.

Let us notice that for the distribution of the lifetime is Lognormal, then Y is a Gaussian random variables of mean μ and variance σ^2 .

Next consider an ALT framework $(Y_i; \mathbf{x}_i)_{1 \leq i \leq n}$ where $\mathbf{x}_i = (x_i^1, \dots, x_i^r)'$ is a vector composed with transformations of (3.2) and Y_i is a random variable of the log lifetime associated with the stress covariable \mathbf{x}_i .

Let $(\hat{a}_0, \hat{a}_1, \dots, \hat{a}_r, \hat{\sigma})$ denote the estimator of the parameter $(a_0, a_1, \dots, a_r, \sigma)$. We estimate the reliability yields under the operating environment - denoting $\mathbf{x}_0 = (x_0^1, \dots, x_0^r)$ - by using $\hat{\mu}(\mathbf{x}_0) = \hat{a}_0 + \sum_{j=1}^r \hat{a}_j x_0^j$ the estimator of $\mu(\mathbf{x}_0)$.

In a recent paper, Meeker and Escobar [26] have shown that ALT leads to clear disagreement with the real comportment of the lifetime when basics assumptions of ALT are not true. They have discussed about the problem of overstressed data which is interesting us for our study. Let us remember that the life-stress relationship (3.2) is valid for some specific stress set. When the stress levels of ALT are going beyond the stress set of the fail; then the experiment reveals several fails, denoting F_1, F_2

, $F_3 \dots$ associated with some stress sets $\mathcal{X}_1, \mathcal{X}_2, \mathcal{X}_3 \dots$ as $\cup_{(1 \leq j)} \mathcal{X}_j = [0; 1]^r$. The fail F_1 is the one which is observed under the operating environmental conditions and the others ones $F_2, F_3 \dots$ are called overstress fails. For overstress data, our aim consists of finding the stress set \mathcal{X}_1 that is associated with the fail F_1 and then the overstress sets $F_2, F_3 \dots$. After, we have to estimate the ALT model defined by (3.1) and the life-stress relationship μ_k defined by (3.2) for the fail F_k for all $k \geq 1$.

As a consequence, the piecewise life-stress relationship is a natural model for overstress ALT data. So we replace (3.2) by the following ones :

$$f(\mathbf{x}) = \sum_{J \in M} \left(a_0 + \sum_{j=1}^{r-1} a_j x^j \right) 1_{(\mathbf{x} \in J)} \quad (3.3)$$

where M is a partition of \mathcal{X} the stress set (associating with the stress sets of the operational fail and overstressed fails). We propose to use a penalized least squares criterion in order to get the stress set of the common fail and of the overstressed fails.

We have considered the least squares estimator (LSE), because this estimation always have an analytic solution (except when the projection matrix is not reversible). And the LSE does not depend on the distribution of the errors ε . Otherwise, the MLE have no analytic solution when the distributions of the lifetimes are Weibull. And the likelihood equations clearly depend on the distribution of ε . Furthermore, thanks to the selection method of Baron, Birgé and Massart [5]; many authors have used their general method for various models, like Sauv e [31, 32]; who have proposed a general penalty function that provides penalized least squares estimation (PLSE) for piecewise polynomial regression and when the error ε as exponential moments around zero (that is the case of Lognormal and Weibull ALT). Thanks to the result of Sauv e, we have a theoretical justification for playing PLSE for overstress ALT data; and we have implemented and calibrated the PLSE when the distribution of the lifetime is Weibull.

3.2 Assumptions & Notations

For simplicity, let us consider the stress set \mathcal{X} as $[0, 1]^r$. And Let $(Y_i, \mathbf{x}_i)_{1 \leq i \leq n}$ be a regression framework where $\mathbf{x}_i = (x_i^1, \dots, x_i^r) \in [0, 1]^r$ is a vector of covariables associated with the i^{st} observation Y_i . The observations are the logarithm of lifetimes from ALT, so we assume that :

$$Y_i = f(\mathbf{x}_i) + \sigma \varepsilon_i, \quad (3.4)$$

where $(\varepsilon_i)_{1 \leq i \leq n}$ are independent and identically distributed random variables of the errors, with mean 0, variance 1 and they have finite exponential moment around 0. After, μ is a function of the stress covariables and σ is non negative constant. We also consider that σ is a known in a first time.

Let \mathcal{M}_n be a collection of partitions of $[0, 1]^r$. For any partitions M into \mathcal{M}_n we define the linear set \mathcal{F}_M of piecewise linear function on M , namely:

$$\mathcal{F}_M = \left\{ \mathbf{x} = (x^1, \dots, x^r) \rightarrow \sum_{J \in M} \left[a_{0,J} + \sum_{j=1}^{r-1} a_{j,J} x^j \right] \cdot 1_{(\mathbf{x} \in J)} : a_{j,J} \in \mathbb{R} \right\}. \quad (3.5)$$

The dimension of \mathcal{F}_M is $D_M = r \cdot |M|$, where $|M|$ is the number of distinct subsets (not commonplace) which compose the partition M .

For a given partition M , let \hat{f}_M denote the LSE of f over \mathcal{F}_M , it is defined as follows :

$$\hat{f}_M = \arg \min_{u \in \mathcal{F}_M} \|Y - u\|_n^2 \quad (3.6)$$

where $\|\cdot\|_n$ is the Euclidian norm of \mathbb{R}^n scaled by a factor $n^{-\frac{1}{2}}$. In (3.6) we have presumed on the notation and when we talk about $\|u\|_n$ for any function u , we use the norm $\|\cdot\|_n$ upon the vector $(u(\mathbf{x}_1), \dots, u(\mathbf{x}_n)) \in \mathbb{R}^n$.

Now for the collection of acceleration functions $(\mathcal{F}_M)_{M \in \mathcal{M}_n}$, we have a collection of candidates $(\hat{f}_M)_{M \in \mathcal{M}_n}$. During this study, we do not consider that f belongs to any piecewise model \mathcal{S}_M . Because we assume that the collection $(\mathcal{F}_M)_{M \in \mathcal{M}_n}$ attempts to approximate the reality. Now for a given partition $M \in \mathcal{M}_n$, we could compare f and \hat{f}_M by using the quadratic risk of the estimator; it decomposes on two terms :

$$\mathbb{E}(\|f - \hat{f}_M\|_n^2) = \|f - f_M\|_n^2 + \sigma \frac{\dim_{\mathbb{R}^n} \mathcal{F}_M}{n} \leq \|f - f_M\|_n^2 + \sigma r \frac{|M|}{n}$$

where f_M is the orthogonal projection of f over \mathcal{F}_M under the norm $\|\cdot\|_n$; then $f_M = \arg \min_{u \in \mathcal{F}_M} \|f - u\|_n^2$. The previous inequality becomes equality when the covariables of the regression frameworks are not collinear. Furthermore, the term $\|f - f_M\|_n^2$ is a bias term; and the term $\sigma \frac{r|M|}{n}$ is a variance term.

Otherwise, we would like to treat many practical cases of overstress ALT data. In many application cases, the errors $(Z_i)_{1 \leq i \leq n}$ shall not always be Gaussian. And as Weibull ALT induces that the errors $(Z_i)_{1 \leq i \leq n}$ are standardized Gumbel random variables, we need an assumption that regroups a larger kind of distribution. So we consider - as Sauvé [31, 32] does - that the errors have exponential moments around 0. So there exists two positive constants τ and b such that :

$$\forall \lambda \in \left(-\frac{1}{b}, \frac{1}{b}\right) \quad \log \mathbb{E}\left(e^{\lambda \sigma \varepsilon_i}\right) \leq \frac{\tau^2 \sigma^2}{2(1 - b|\lambda|)} \quad (3.7)$$

Let us notice that when ε_i are standardized Gumbel random variables, the inequality (3.7) holds with $\tau = \sigma \sqrt{2\gamma}$ and $b = \frac{2}{\sigma}$. Generally τ^2 is larger than σ^2 and we could choose it as close as we want to σ^2 but at the price that b is larger.

Next we suppose that M_0 is a fine partition of $[0, 1]^r$. And we consider that \mathcal{M}_n is a set of partitions M where all disjoint subsets J defining M are an union of elements of M_0 . In a few words, the fine partition M_0 acts as a canonical basis of all partitions; then we define \mathcal{M}_n as the set of the exhaustive partitions build from M_0 . For the numerical applications, it is important that $N_n = |M_0|$ is large enough. Otherwise the test plan has to be consistent with the choice of M_0 . For a given $J_0 \in M_0$, we define $|J_0|$ as, $|J_0| = \#\{\mathbf{x}_i : \mathbf{x}_i \in J_0 \text{ and } 1 \leq i \leq n\}$. If M_0 is too large, the estimation of \hat{f}_{M_0} could lead to disagreements. As a consequence, the value of $N_{\min} = \arg \min_{J_0 \in M_0} |J_0|$ have to be large enough in practice. The selection theorem on the next section gives us a theoretical minimum value for N_{\min} .

3.3 Selection theorem

A classical method of model selection consists of building a penalized selection criterion. Let pen be a penalty function which takes values from \mathcal{M}_n to $\mathbb{R}+$ and it increases with $|M|$. The penalized criterion selects \hat{M} the «*best*» partition among \mathcal{M}_n , namely :

$$\hat{M} = \arg \min_{M \in \mathcal{M}_n} \{ \|Y - \hat{f}_M\|_n^2 + pen(M) \}.$$

The ideal model, indexed by M^* , would be the one which is minimizing the risk $\mathbb{E}(\|f - \hat{f}_M\|_n^2)$ over all $M \in \mathcal{M}_n$. Unfortunately the oracle partition M^* is unknown since f and f_M are unknown too. Thanks to the bias/variance decomposition of the risk :

$$\mathbb{E}(\|f - \hat{f}_M\|_n^2) = \|f - f_M\|_n^2 + \sigma \frac{\dim_{\mathbb{R}^n} \mathcal{F}_M}{n} \leq \|f - f_M\|_n^2 + \sigma \frac{r|M|}{n}.$$

Then a «*good*» estimation of M need to do a compromise between the bias term and the variance term. Referring to the LSE loss function, we minimize the variability of the estimator and we clearly encourage larger models. So here a «*good*» criterion shall penalize the larger models. The penalized selection criterion attempts to mimic the comportment of the oracle M^* by using a penalty function.

The classical selection penalty are based on asymptotic considerations : as an example see Mallows' C_p [22], Akaike's AIC [1] or Schwarz's BIC [33]. However when the models are large, the values of $\|\hat{f}_M\|_n^2$ could be far from its expectation. And then the classical penalty may lead to underestimate the bias between f and \hat{f}_M . More recently Baron, Birgé and Massart [5] have proposed to use concentration inequalities in order to build penalized selection methods.

Let $\theta \in (0, 1)$ be a real, let M be a partition of $[0, 1]^r$ and defining f_M (resp. Z_M) as the orthogonal projection of f (resp. $Z = \sigma\varepsilon$) over \mathcal{F}_M using the norm $\|\cdot\|_n$, we have $f_M = \arg \min_{u \in \mathcal{F}_M} \|f - u\|_n^2$ (resp. $Z_M = \arg \min_{u \in \mathcal{F}_M} \|Z - u\|_n^2$). Thanks to Pythagore's inequality and the definitions of \hat{f}_M and \hat{M} , we have :

$$\begin{aligned} (1 - \theta) \|f - \hat{f}_{\hat{M}}\|_n^2 &= (2 - \theta) \|Z_{\hat{M}}\|_n^2 - 2 \langle Z, f - \hat{f}_{\hat{M}} \rangle_n - \theta \|f - \hat{f}_{\hat{M}}\|_n^2 - pen(\hat{M}) + \\ &+ \inf_{M \in \mathcal{M}_n} \{ (2 - \theta) \|Z_M\|_n^2 - 2 \langle Z, f - \hat{f}_M \rangle_n + \theta \|f - \hat{f}_M\|_n^2 - \\ &- pen(M) \}. \end{aligned}$$

The mean idea of Baron, Birgé and Massart [5] consists of using concentration inequalities in order to control the following terms :

- $\langle Z, f - \hat{f}_M \rangle_n^2$: by using the concentration inequality of a sum of random variables (as Log Laplace inequality [21]);
- $\chi_M^2 = \|Z_M\|_n^2$: by using the concentration inequality of Bernstein [21] (or Cirelson, Ibragimov and Sudakov's inequality when ε is gaussian) on a truncated version of χ_M^2 .

Lastly the selection theorem gives us the form of the penalty function that ensures that the PLSE exists, is consistent and we exhibit an oracle inequality. Unfortunately the results of Birgé and Massart [6] are not valid for non Gaussian linear models. And here we consults the selection theorem due to Sauvé [31] that extends the result of Birgé and al. to non Gaussian piecewise polynomial models. Next we have adapted the theorem of Sauvé for piecewise ALT models (3.3), then the selection theorem is:

Theorem 3.1 (Sauvé, 2006). *Let $b \in \mathbb{R}^+$ and $\tau \in \mathbb{R}^*+$ such that inequality (3.7) holds.*

Assume the points $(\mathbf{x}_i)_{1 \leq i \leq n}$ to be distributed such that :

$$\|F_n - F\|_\infty \leq \frac{24A(r)^2}{2^r(4B(r) + 1)} \frac{b^2}{\tau^2} \frac{\log n}{n} \quad (3.8)$$

where F_n is the empirical distribution function associated with $(\mathbf{x}_i)_{1 \leq i \leq n}$, F is the uniform distribution function on $[0, 1]^r$, and $\|\cdot\|_\infty$ is the supnorm on $[0, 1]^r$ and $A(r)$ and $B(r)$ are theoretical constants which come from equivalence of the norms.

Let M_0 a partition of $[0, 1]^r$ composed of hyper rectangle axis oriented regions and assume that $N_{\min} = \inf_{J \in M_0} |J|$ satisfies :

$$N_{\min} \geq 24A(r)^2 \frac{b^2}{\tau^2} \log n$$

Let M_n a family of partitions of $[0, 1]^r$ composed of hyper rectangle axis oriented regions built from those of M_0 , and $(w_M)_{M \in \mathcal{M}_n}$ a family of weights such that :

$$\sum_{M \in \mathcal{M}_n} e^{-w_M} \leq \Sigma \in \mathbb{R}^+ \quad (3.9)$$

Assume $\|f\|_\infty \leq R$, with R a positive constant.

Let $\theta \in (0, 1)$ and $K > 2 - \theta$ two numbers.

Taking a penalty satisfying :

$$\begin{aligned} \text{pen}(M) \geq & K.r \frac{\tau^2}{n} |M| + 8\sqrt{2}(2 - \theta)r \frac{\tau^2}{n} \sqrt{r|M|w_M} \\ & + \left[\left(4(2 - \theta)r + \frac{2}{\theta}\right) \frac{\tau^2}{n} + 2A'(r) \frac{Rb}{n} \right] w_M \end{aligned} \quad (3.10)$$

we have :

$$\begin{aligned} \mathbb{E}(\|f - \hat{f}_M\|_n^2) \leq & \frac{2}{1 - \theta} \inf_M \{ \|f - f_M\|_n^2 + \text{pen}(M) \} \\ & + \frac{1}{1 - \theta} \left[8(2 - \theta)r \left(1 + 8r \frac{(2 - \theta)}{K + \theta - 2} \right) \frac{4}{\theta} + 2 \right] \frac{\tau^2}{n} \Sigma + \\ & + \frac{6A'(r)}{1 - \theta} \frac{Rb}{n} \Sigma + C(b, \tau^2, R, r) \frac{1_{(b \neq 0)}}{n(\log n)^{\frac{3}{2}}} \end{aligned}$$

where $C(b, \tau^2, R, r)$ is a positive constant which only depends on b , τ^2 , R , r , and $A'(r) = 1 + \sqrt{2}rA(r)$.

Notices that the theorem makes an additional assumption that the empirical distribution of the covariables $(\mathbf{x}_i)_{1 \leq i \leq n}$: the distance between F_n empirical distribution of the covariables and F the uniform distribution on $[0, 1]^r$ has to be bounded as follows :

$$\|F_n - F\|_\infty \leq \frac{24A(r)^2}{2^r(4B(r) + 1)} \frac{b^2 \log n}{\tau^2 n} \quad (3.8)$$

where the constants $A(r)$ and $B(r)$ are universal constants that depend on the canonical basis using during the piecewise estimation. More precisely, the constants $A(r)$ and $B(r)$ ensure the equivalence between the norm-sup of $L^\infty([0, 1]^r)$ and the discrete norms $\|\cdot\|_1$ and $\|\cdot\|_2$ where for u a function of $[0, 1]^r$ $\|u\|_1 = \sum_{i=1}^n |u(\mathbf{x}_i)|$ and $\|u\|_2^2 = \sum_{i=1}^n u^2(\mathbf{x}_i)$ (see Sauvé [31]), lemma 3.4 and 3.5 for more details).

Remark 3.3.1. This assumption (3.8) is an elegant hypothesis which ensures that all $\|\phi_{k,J}\|_{|J|}^2$ are bound where J is included into the given partition M and the family $(\phi_{k,J})_k$ is an orthonormal basis of the life-stress relationship. As an example for Arrhenius acceleration, $f(x) = a_0 + a_1 \cdot x$ with $(\mathbf{x}_i)_{i=1, \dots, n}$ are the inverse of the temperatures, $\phi_{0,J}(x) = \frac{1_{(x \in J)}}{|J|}$ and $\phi_{1,J}(x) = \frac{(x - \bar{\mathbf{x}}_J) 1_{(x \in J)}}{\|(\mathbf{x} - \bar{\mathbf{x}}_J) 1_{(\mathbf{x} \in J)}\|_{|J|}}$ where $|J| = \sum_{i=1}^n 1_{(\mathbf{x}_i \in J)}$ and $\bar{\mathbf{x}}_J = \sum_{i=1}^n \mathbf{x}_i \cdot 1_{(\mathbf{x}_i \in J)}$. Here the $\|\phi_{k,J}\|_{|J|}^2$ are bounded since we assume that both $\min_{J \in M} |J|$ and the ratio between the bias and the variance of $(\mathbf{x}_i)_{i=1, \dots, n}$ upon J are bounded.

Briefly the covariables using for the test plan do not need to be uniformly distributed. Every subsets of the fine partition M_0 have to include neither too many nor enough covariables points $(\mathbf{x}_i)_{i=1, \dots, n}$. Let us remark that a test plan of ALT usually recommends concentrating the observations under the higher stress levels. This *ad-hoc* recommendation is compatible with the assumption (3.8) about the distribution of $(\mathbf{x}_i)_{i=1, \dots, n}$: for Arrhenius life-stress relationship, we only have to take care that the ratio between the internal bias and the variance over the $(\mathbf{x}_i)_{i=1, \dots, n}$ of all $J \in M_0$ is bounded. And the more F_n is close to F , the larger M_0 could be chosen.

Moreover this theorem gives the general form of the penalty function :

$$pen(M) = K.r. \frac{\tau^2}{n} |M| + \left(\kappa_1(\theta, r) \frac{\tau^2}{n} \sqrt{|M| w_M} + \kappa_2(\theta, r) \frac{\tau^2 + Rb}{n} w_M \right) \quad (3.11)$$

The penalty function could be seen as the sum of two terms. Firstly we have a variance term that is proportional with the variance term of the previous decomposition of the quadratic risk of \hat{f}_M . And secondly we have a complexity term that depends on the family of weights $(w_M)_{M \in \mathcal{M}_n}$. We often tell that the weights $(w_M)_{M \in \mathcal{M}_n}$ are defining a kind of *posterior* probability law of partitions $M \in \mathcal{M}_n$ candidates to the estimation, as we could always choose (w_M) such that the sum $\sum_M e^{-w_M}$ is 1.

The selection theorem ensure that for $\theta \in (0, 1)$ and $K > 2 - \theta$, the PLSE $\hat{f}_{\hat{M}}$ exists, it is consistent and satisfies the following oracle inequality:

$$\mathbb{E} \left(\|f - \hat{f}_{\hat{M}}\|_n^2 \right) \leq C_1 \cdot \inf_{M \in \mathcal{M}_n} \{ \|f - f_M\|_n^2 + pen(M) \} + \frac{C_2}{n}$$

where C_1 depends on θ and C_2 depends on f (through R), on r , on weights (through Σ) and on the distribution of the errors $\varepsilon = \sigma \cdot Z$ (through b and τ). Since the

oracle inequality holds, we have the theoretical proof that the PLSE would mimic the comportment of the oracle model : so we valid the selection method.

For overstress ALT data, we have to estimate a piecewise acceleration function. Let $(Y_i, \mathbf{x}_i)_{1 \leq i \leq n}$ be an ALT framework and considers M_0 as a fine partition of the stress covariable set $\{\mathbf{x}_i : i = 1, \dots, n\} \subseteq \mathcal{X}$ an hyper rectangle of dimension r and $N_n = |M_0|$. Since we do not have any *a priori* information about f , we have to consider \mathcal{M}_n as the collection of all partitions of \mathcal{X} using the elements of M_0 as an canonical basis of partitions. We usually define $(w_M)_{M \in \mathcal{M}_n}$ such that w_M only depends on M throught its dimension $|M|$ (see in previous section 2.3.4 or [6, 19, 32] for more details). Since $1 \leq k \leq N_n$, $\#\{M \in \mathcal{M}_n : k = |M|\} = \binom{N_n-1}{k-1}$ and if we define $w_M = |M|(L + \log(\frac{N_n}{|M|}))$ with $L > 1$, than the sum $\sum_M e^{-w_M} \leq (e^{L-1} - 1)^{-1}$. Now applying the selection theorem, we assume the oracle type inequality holds since we use the following penalty function :

$$pen(M) = (\tau^2 + bR) \frac{r \cdot |M|}{n} \left[c_1 \log \left(\frac{N_n}{|M|} \right) + c_2 \right] \quad (3.12)$$

where the constants c_1 and c_2 are absolute calibration constants, they only depends on the thinness of the bound used during the theorem and its application. We do not give any values for (c_1, c_2) (such as we did for K and θ), because they are theoretical constants. Birgé and Massart [7] and Massart [21] have studied the consistency and the convergence rate of the PLSE for several the scale of the calibration constants for Gaussian data. Furthermore, we believe that in numerical application we could get optimal values of the calibration pair (c_1, c_2) .

Lastly we propose a selection procedure in order to selection piecewise model when the data $(Y_i, \mathbf{x}_i)_{1 \leq i \leq n}$ are overstressed. The procedure runs on two steps :

- Firstly, calculates for $k = 1, \dots, N_n$ the « best » partitions \hat{M}_k on k pieces:

$$\hat{M}_k = \arg \min_{M \in \mathcal{M}_n : |M|=k} \|Y - \hat{f}_M\|_n^2$$

see appendix A for more details about the research of partitions ;

- Secondly, thanks to the pen function (3.12) and denotes $f_k = f_{\hat{M}_k}$, we estimate \hat{k} the number of pieces, namely :

$$\hat{k} = \arg \min_{k=1, \dots, n} \{ \|Y - \hat{f}_{\hat{M}_k}\|_n^2 + pen(\hat{M}_k) \}.$$

At the end we get $\tilde{M} = \hat{M}_{\hat{k}}$ and $\tilde{f} = \hat{f}_{\hat{k}}$ in order to define the last estimation of M and f .

3.4 First method

Thanks to the selection theorem, we assume that the general form of the penalty function is :

$$pen(M) = \frac{|M|}{n} \left(\alpha \log \left(\frac{N_n}{|M|} \right) + \beta \right)$$

with two calibration constants α and β which have to be calibrated. We proposed a first method of calibration based on the data and using Birgé-Massart's heuristic (see [21], section 8.5.2). In a few words, the heuristic method proposes to fit the form of the penalty on $-||Y - \hat{f}_{\hat{M}_k}||_n^2$ when k is large. And $(\hat{\alpha}, \hat{\beta})$ are estimating to fit $-||Y - \hat{f}_{\hat{M}_k}||_n^2$ on $\frac{1}{2} pen(\hat{M}_k) + \gamma$ when k is large. Let us notice that such a writing of the penalty is including the shape variable σ and R . So the method shall also be using when σ and R are unknown.

Firstly we have recalled the principle of the heuristic. Secondly we have simulated the heuristic for an overstress ALT data when we have 3 overstress sets, Arrhenius life-stress relationship and the distribution of the lifetime is Weibull.

3.4.1 Birgé-Massart's heuristic

Birgé-Massart's heuristic could be seen as a generalization of Mallows' heuristic (see [21], [31] or section 2.5.1). Mallows [22] have proposed a penalized criterion for Gaussian model selection using LSE. Thanks to the Cochran's theorem, the quadratic risk decomposes as follows :

$$\mathbb{E}(\|f - \hat{f}_M\|_n^2) = \|f - f_M\|_n^2 + \sigma^2 r \frac{|M|}{n}.$$

So the model that minimizes the quadratic risk in the one which minimizes :

$$-\|f_M\|_n^2 + \sigma^2 r \frac{|M|}{n}.$$

As $\|f_M\|_n^2$ is unknown, Mallows has estimated it by an unbiased estimator $\|\hat{f}_M\|_n^2 - \sigma^2 r \frac{|M|}{n}$. Lastly Mallows proposes to minimize the following criterion :

$$C_p(M) = \frac{1}{n} \sum_{i=1}^n [Y_i - \hat{f}_M(\mathbf{x}_i)]^2 + 2\sigma^2 r \frac{|M|}{n}.$$

On a first hand, we know that the random term $\|\hat{f}_M\|_n^2$ (resp. $\|Y - \hat{f}_M\|_n^2$) have not always the same order of magnitude as its expectation for all partition M simultaneously. As an example, since the partition M is large, the value of $\|\hat{f}_M\|_n^2$ (resp. $\|Y - \hat{f}_M\|_n^2$) does not keep on the same order of magnitude as its expectation. When M is a fine partition of the stress set, we could not valid the heuristic of Mallows.

On the other hand, when M is large enough, we assume that the bias between $\|Y - \hat{f}_M\|_n^2$ and its expectation is constant and is free itself from M (see [21], [31] or section 2.5.1). Mallows has proposed to get the following penalty : $pen(M) = 2\sigma^2 r \frac{|M|}{n}$. And the bias correction of $\|Y - \hat{f}_M\|_n^2$ is half of the penalty function. By analogy with Mallows' heuristic, Birgé-Massart's heuristic assumes that the penalty function pen have to compensate the half of $\|Y - \hat{f}_M\|_n^2$ when the partition M is a large partition. But the new penalty function takes account of the model's

complexity. At the end, referring to the notations used in section 3, the heuristic consists of fitting the calibration pair (α, β) as follows :

$$\|Y - \hat{f}_k\|_n^2 \simeq \frac{1}{2} \frac{k}{n} \left[\alpha \log \left(\frac{N_n}{k} \right) + \beta \right] + \gamma$$

where $k_{\min} \leq k \leq k_{\max}$.

Remark 3.4.1. Notices that the τ^2 , b and R do not appear into the writing of the penalty function. We have integrated them into the calibration pair (α, β) . Because τ^2 , b and R are technical constants that characterize the form of the penalty function. So as soon as we have an approximation of the penalty function, we do not need any approximation or estimation of τ^2 , b and R .

3.4.2 Numerical application

In order to test the first selection method, we have simulated overstress ALT data for various sampling sizes. We used, as reference model a piecewise ALT model in 3 pieces, with Weibull distribution and Arrhenius life-stress relationship. As we are working on log-lifetimes, we have simulated samples through Gumbel's distribution. We have simulated 100 log-lifetimes using a scale parameter $\sigma = 0.8$ and the following piecewise acceleration function:

$$f(x) = \begin{cases} -24 + 12900/x, & \text{if } x < 345; \\ -145 + 52900/x, & \text{if } 345 \leq x < 370; \\ -46 + 13400/x, & \text{if } 370 \leq x. \end{cases}$$

The sampling sizes are $n=50, 100, 200$ and 400 . And the covariables are fixed as they are a regular grid in n pieces of the interval $[300, 400]$. The figure 3.1 represents an acceleration model using for simulations. The red line is f and the blue one crosses are the simulating log-lifetimes.

The aim of this section is to show the first method at work. So we have implemented the method when n varies and for many values of k_{\min} and k_{\max} , in order to show what magnitude of k_{\min} and k_{\max} leads to *good* results. Also we have compared the distribution of \hat{c} and $\hat{\lambda}$ when n, k_{\min} and k_{\max} varies. As we have worked on simulating data, we are able to compare the ratio between the risk of the PLSE and the oracle risk. Then we could valid that the heuristic of Birgé-Massart has the same goal as the selection theorem of section 3 : mimicking the oracle model. So lastly, we have compared the risk ratios and the distribution of \hat{k} in order to evaluate the performances of the method.

The tables 3.2 and 3.3 and the figures 3.2 and 3.3 summarize the Monte-Carlo estimations of c and λ using the heuristic of Birgé-Massart. During the simulated study, we have simulated 100 ALT framework and for the j^{st} sample let $(\hat{c}^{(j)}, \hat{\lambda}^{(j)})$ be the estimation of the calibration pair (c, λ) based on Massart's heuristic for fixed values of k_{\min} and k_{\max} . The table 3.2 shows the median of $\hat{c}^{(j)}$ and $\hat{\lambda}^{(j)}$ where $j = 1, \dots, 100$. And it also shows an estimation of the risk ratio between the risk of the PLSE and the risk of the oracle models. The table 3.3 presents the empirical distribution of the PLSE

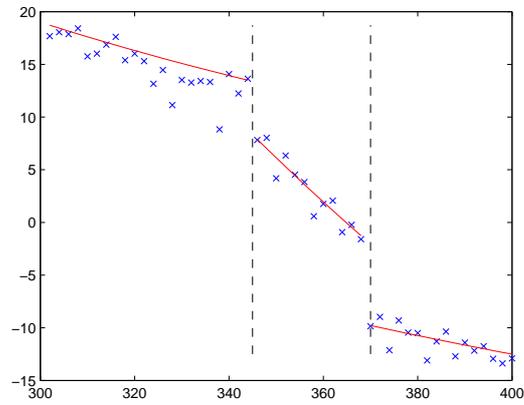


Figure 3.1: The plot is drawing (red line) the 3-piecewise acceleration function using for the simulations, and we have also drawn (blue crosses) the log-lifetimes of a simulating sample using Gumbel's distribution.

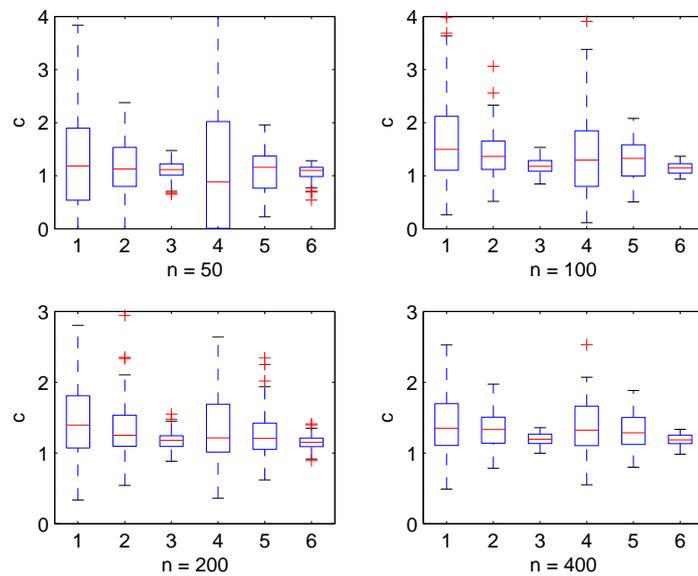


Figure 3.2: Boxplot for c when the lifetime distribution is Weibull and various series of sampling sizes, in $1 \leq k \leq n/4$, in $2 \leq k \leq n/3$, ... and n varies from 50 to 400.

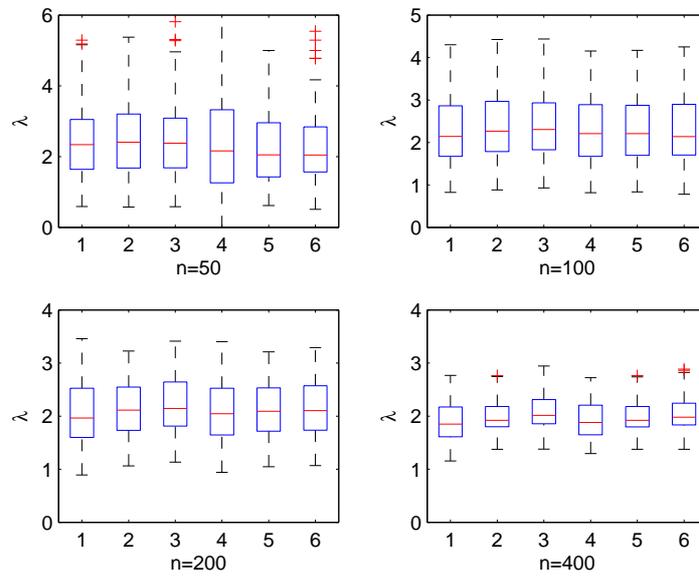


Figure 3.3: Boxplot for λ when the lifetime distribution is Weibull and various series of sampling sizes, in 1 $5 \leq k \leq n/4$, in 2 $5 \leq k \leq n/3$, ... and n varies from 50 to 400.

(k_{\min}, k_{\max})	$(5, \frac{n}{4})$	$(5, \frac{n}{3})$	$(5, \frac{n}{2})$	$(10, \frac{n}{4})$	$(10, \frac{n}{3})$	$(10, \frac{n}{2})$
$n=50: \hat{c}$	1.0879	1.1654	1.1062	1.5759	1.0996	1.0722
$\hat{\lambda}$	2.3392	2.4069	2.3798	2.1589	2.0485	2.0447
risk ratio	2.1354	2.2325	1.4781	5.3012	1.9602	1.6843
$n=100: \hat{c}$	1.4536	1.3529	1.1972	1.4142	1.2655	1.1492
$\hat{\lambda}$	2.1426	2.2670	2.3032	2.2061	2.2082	2.1380
risk ratio	1.0000	1.0000	1.0000	1.0000	1.0026	1.0059
$n=200: \hat{c}$	1.4532	1.3450	1.2110	1.3217	1.2898	1.1920
$\hat{\lambda}$	1.9650	2.1120	2.1437	2.0450	2.0913	2.1040
risk ratio	1.0025	1.0000	1.0000	1.0025	1.0000	1.0025
$n=400: \hat{c}$	1.4814	1.3639	1.1968	1.4218	1.3238	1.1810
$\hat{\lambda}$	1.8491	1.9219	2.0138	1.8796	1.9215	1.9790
risk ratio	1.0196	1.0000	1.0000	1.0000	1.0000	1.0000

Table 3.2: The table represents the median estimations of c^* and λ^* and the empirical risk ratio for Weibull ALT with various sampling sizes.

(k_{\min}, k_{\max})	$(5, \frac{n}{4})$	$(5, \frac{n}{3})$	$(5, \frac{n}{2})$	$(10, \frac{n}{4})$	$(10, \frac{n}{3})$	$(10, \frac{n}{2})$
$n=50: k$	2 3 4 5	2 3 4 5	2 3 4 5	2 3 4 5	2 3 4 5	2 3 4 5
$\hat{p}_{\hat{k}}(k)$	6 90 0 1	7 90 0 0	4 95 0 1	4 67 0 0	6 77 1 2	5 89 1 0
$n=100: k$	2 3 4 5	2 3 4 5	2 3 4 5	2 3 4 5	2 3 4 5	2 3 4 5
$\hat{p}_{\hat{k}}(k)$	0 100 0 0	0 100 0 0	0 100 0 0	0 100 0 0	0 99 1 0	0 98 2 0
$n=200: k$	2 3 4 5	2 3 4 5	2 3 4 5	2 3 4 5	2 3 4 5	2 3 4 5
$\hat{p}_{\hat{k}}(k)$	0 99 1 0	0 100 0 0	0 100 0 0	0 99 1 0	0 100 0 0	0 99 1 0
$n=400: k$	2 3 4 5	2 3 4 5	2 3 4 5	2 3 4 5	2 3 4 5	2 3 4 5
$\hat{p}_{\hat{k}}(k)$	0 98 1 0	0 100 0 0	0 100 0 0	0 100 0 0	0 100 0 0	0 100 0 0

Table 3.3: The table represents the empirical distribution of k using the PLSE for Weibull ALT with various sampling sizes.

of k the number of acceleration. And the figure 3.2 (resp. 3.3) shows us the box plot of $(\hat{c}^{(j)})_{j=1, \dots, 100}$ (resp. $(\hat{\lambda}^{(j)})_{j=1, \dots, 100}$).

Thanks to the figure 3.2 we see that when k_{\max} is increasing, the variability of the estimation of c is decreasing. And for a given k_{\min} , the box plots are nested when k_{\max} increases. Otherwise, the calibration constant λ does not seem to be affected by the values of k_{\min} or k_{\max} .

And looking at the table 3.2, we see that the risk ratio between the risk of the estimated PLSE and the oracle risk is almost 1, except when $n=50$, then the risk ratio is decreasing when k_{\max} is increasing. And, the table 3.3 tell us that the method clearly finds the real number of acceleration since $n \geq 100$. And when $n = 50$, the method also works since k_{\min} is not far from the number of ruptures.

Lastly, the heuristic of Birgé-Massart is a practical rule which builds a data-driven penalized criteria and it is mimicking the oracle model when n is large enough, k_{\min} is not far from the number of ruptures and k_{\max} does not need to be too large.

3.5 Second method

Thanks to the selection theorem, the general form of the penalty is :

$$\text{pen}(M) = (\tau^2 + bR) \frac{r \cdot |M|}{n} \left[c_1 \log \left(\frac{N_n}{|M|} \right) + c_2 \right]$$

where c_1 and c_2 are calibration constants.

We could interpret this general penalty function as the sum of two terms :

- variance term : $c_2(\tau^2 + bR) \frac{r \cdot |M|}{n}$;
- complexity term : $c_1(\tau^2 + bR) \frac{r \cdot |M|}{n} \log\left(\frac{N_n}{|M|}\right)$.

The variance term depends on the distribution the error (Gaussian, Gumbel, ...) and on the kind of model (as histogram, piecewise linear regression, ...). And the complexity term depends on the number of models for a given dimension, $\#\{M \in \mathcal{M}_n : |M| = k\}$. If we use data from the same family of models, than the suitable value of the variance term *a priori* does not change. And if we use the same research algorithm to find the partitions candidates \mathcal{M}_n , then the complexity term *a priori* does not change. Lastly, we notices that the scale of variation of the penalty funtion also depends on σ (via τ and b) and μ (via R).

Based on these statements, we present here a second method of calibration. This alternative method consists of calibrating the penalty for a given family of models. That kind of method was already used by Birgé and Rozenholc [8] and Lebarbier [19]. Firstly, we use a Monte Carlo study in order to estimate c^* , the ratio between the variance and the complexity term of the penalty as the following penalty function which is built in order to mimic the oracle model :

$$pen(M) = \lambda \frac{|M|}{n} \left[\log\left(\frac{N_n}{|M|}\right) + c^* \right].$$

Secondly, we use the form of the penalty given by the MC-study and we apply the heuristic of Birgé and Massart [6] in order to correct the order of magnitude λ^* of the penalty through the data. Lastly, we get a data-driven penalty, namely :

$$pen(M) = \lambda^* \frac{|M|}{n} A(M);$$

with $A(M) = \left[\log\left(\frac{N_n}{|M|}\right) + c^* \right]$.

3.5.1 Estimation of c^*

The procedure of Birgé and al. [8] consists of studying the ratio between the risk of the PLSE and the oracle risk. The main idea of the procedure is to choose an «*optimal*» calibration pair (λ, c) . The calibration pair is obtained to be optimal as soon as the PLSE estimator mimics the oracle estimator (see also the proof of the selection theorem [32]). In a way, the procedure is directly based on the theoretical statements of the selection theorem 3.1. So the optimal calibration pair $(\hat{\lambda}, \hat{c})$ is usually chosen as the risk ratio between the risk of the PLSE and the oracle risk is uniformly close to 1 for a given family of models and for a given family of sampling sizes. See as an example the results of Lebarbier ([19], on the chapter 3) who has treated the case of constant piecewise Gaussian models. In that example, the author has estimated the risks through a Monte-Carlo procedure and as she was working with a MC-procedure, the oracle models are well-known. However, we have to notice that the procedure is costly and is not always

easy to interpret (see the works of Lebarbier [19] or the section 2.4.3 on the previous chapter).

Let us notice that the shape factor of the penalty function was σ^2 on the previous application of the calibration procedure of Birgé and Rozenholc. But here the shape factor on the penalty function was $(\tau^2 + bR)$ instead of σ^2 ; and this term depends on the data through R . Moreover, we are able to calculate the constants τ and b for a fixed distribution of the errors. But we are also able to find several constants τ and b which assume that the inequality (3.7) holds. As the shape magnitude of the penalty function *a priori* depends on the data R and an arbitrary choice of (τ, b) ; there is no reasons that the costly MC-procedure of Birgé and Rozenholc works with the new penalty when $b \neq 0$. And we prefer a time procedure that could easily be implementing.

Thus b could be chosen as small as we want at the price that τ is larger than σ . We prefer estimating c through the use of Birgé-Massart's heuristic on full noisy data. And after, we will have to correct the multiplicative factor $\lambda = \sigma^2$ of the penalty function (see the next subsection).

The second method of calibration is based on the use of Birgé-Massart's heuristic for full noisy data. Let us recall that the aim of the first method is to model the penalty function on the half of the loss function of the LSE estimator when the model is large enough. Consider a full noisy data, since $f = 0$ all models should over estimate the real model. In such a case, we would apply the first method with $k_{\min} = 1$. Since we simulate several full noisy data, we could get the distribution of c for a given sampling size n and at a fixed k_{\max} .

Numerical application

In this section, we have implemented the second calibration method of c . We have simulated 100 frameworks of a full noisy data using a standardized version of Gumbel's distribution ($\frac{\pi}{\sqrt{6}}\varepsilon^* \sim \text{SEV}(0, 1) + 0.5772\dots$). We have supposed that $n=50, 100, 200$ and 400 and $k_{\max}=\frac{n}{5}, \frac{n}{4}, \frac{n}{3}$ and $\frac{n}{2}$. We have considered that the model is a piece-wise Arrhenius Weibull model, and we have consider that $\sigma = 1$ and the n -sample of full noisy log lifetimes is associated with a covariable \mathbf{x} is an regular grid of $(n - 1)$ temperatures between 300 and 400°C .

k_{\max}	$n/5$	$n/4$	$n/3$	$n/2$
$n=50$	1.4160	1.5972	1.4030	1.2180
$n=100$	1.6871	1.5504	1.3573	1.1994
$n=200$	1.5392	1.4351	1.3591	1.2067
$n=400$	1.4481	1.3951	1.3287	1.2089

Table 3.4: The table represents the median estimations of c^* based on the pure noisy method for Weibull ALT with various sampling sizes.

The table 3.4 shows us the median of the 100 estimators of c when the sampling size n varies and when the value of k_{\max} using on the heuristically method varies too. We see that the median is decreasing when k_{\max} is increasing (except when $n=50$ and $k_{\max} = \frac{n}{5}$). Moreover, if k_{\max} is increasing, then the variability of the medians (using the sampling sizes) decreases. Such a converging trend is more relevant in the box plots.

On the figure 3.4, we have plotted out the box plots of the estimation of c for some fixed sampling size n and for a fixed value of k_{\max} . When n is fixed, we could see that the box plots are nested and the values of \hat{c} seems to converge at a value $c^* \simeq 1.2$. Also we see that the empirical density of estimations of c is not symmetric and the magnitude of variation of \hat{c} is decreasing when k_{\max} is increasing. And the table 3.4 shows for a given sampling size n and at a fixed k_{\max} . Thanks to the table 3.4 and the box plots 3.4 we assume that for larger values of k_{\max} , the c .

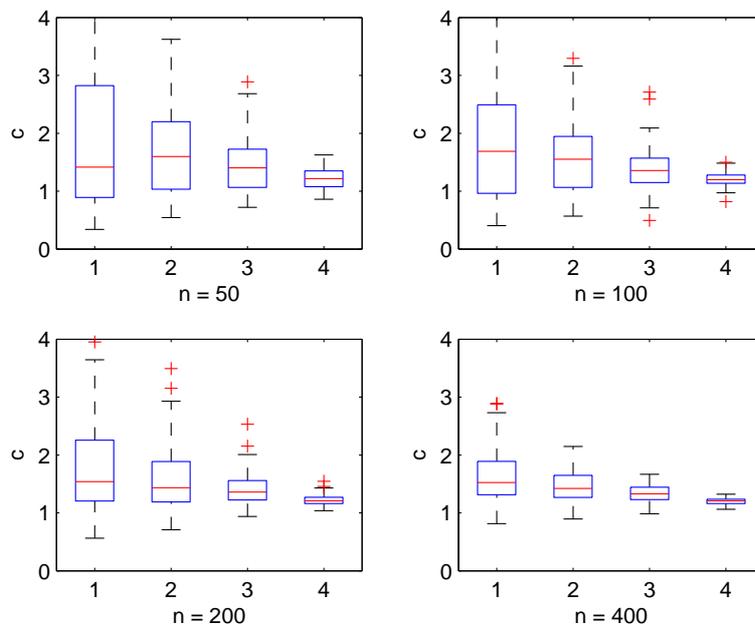


Figure 3.4: Boxplot for c when the lifetime distribution is Weibull and various series of sampling sizes.

3.5.2 Estimation of λ^*

For Gaussian data, Birgé and Massart [6] have showed that a minimal penalty exists, namely :

$$pen_{\min}(M) = \sigma^2 \frac{|M|}{n} g_n(|M|)$$

where $g_n(\cdot)$ is a known form of the penalty function which only depends on the model through its dimension. In details, they have bounded the penalty function. They have proved that :

- When the penalty is namely : $pen(M) = (1 + \eta)\sigma^2 \frac{|M|}{n} g_n(|M|)$, with $\eta \geq 0$; then there is an oracle inequality;
- When the penalty is namely : $pen(M) \geq (1 + \eta)\sigma^2 \frac{|M|}{n} g_n(|M|)$, with $\eta \geq 0$; if the dimension of M is large enough, then the risk of the PLSE can be arbitrarily large.

The authors prefer to use the following heuristic : getting the minimal penalty, as $pen_{\min}(M) = \lambda_{\min} \frac{|M|}{n} g_n(|M|)$; and defining the last penalty such that $pen(M) = 2 pen_{\min}(M)$.

Such an heuristic is useful in many practical cases :

- One would have a data-drive correction of the magnitude of the penalty function, because an under approximation of the optimal penalty function leads to overestimation ;
- One would like to estimate σ , because in many practical cases we are not able to assume that it is not realistic to assume that σ is known.

Recalling that in our general case, we have implemented a calibration method for c^* ; but we do not know λ^* , the magnitude of the penalty function :

$$pen(M) = \lambda \frac{|M|}{n} g_n(|M|)$$

with $\lambda^* \geq \sigma$ and $g_n(k) = \left[\log \left(\frac{N_n}{k} \right) + c^* \right]$.

We propose to apply the heuristic of Birgé and Massart based on the minimal penalty function ; so we would have a data-driven method of calibration of pen (see chapter 2, section 2.6.1). The heuristic is based on the dimensional gap. Considers that $\lambda \geq 0$ the shape of the penalty is fixed, and let \hat{k}_λ denote the dimension of the model selecting by the PLSE associated with λ , we have :

$$\hat{k}_\lambda = \arg \min_{k=1, \dots, N_n} \left\{ \|Y - \hat{f}_k\|_n^2 + \lambda \frac{k}{n} g_n(k) \right\}.$$

If λ the magnitude of the penalty is such that $\lambda \ll \lambda^*$, the dimension \hat{k}_λ should be around N_n (the dimension of the largest model). And if $\lambda_{\text{opt}} \approx \lambda_{\min}$, then we should see a dimensional gap on the plot of the graph $\lambda \rightarrow \hat{k}_\lambda$. By definition, the dimension

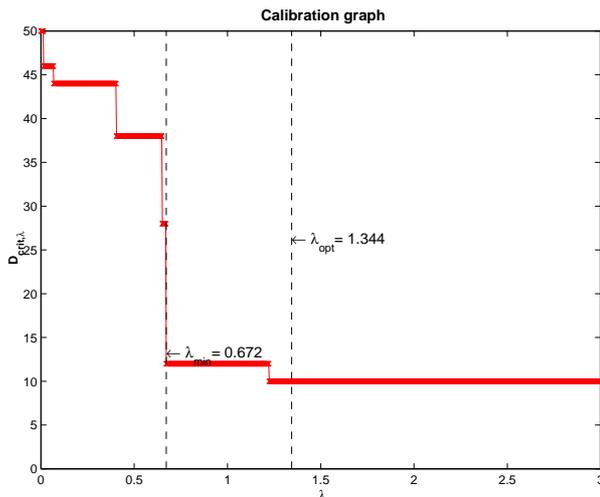


Figure 3.5: The heuristic trend in practice for 5 piecewise Arrhenius Weibull problem.

of the gap is $\hat{k}_{\lambda_{\min}}$. The figure (3.5) shows the application of the heuristically method for 4 piecewise Arrhenius Weibull framework. We observe the gap between the higher dimensions and the penalized dimensions when $\lambda = 0.672$. At last, the method proposes to choose $\lambda_{\text{opt}} = 1.344$ and here the PLSE selects the piecewise model with 5 pieces.

Remark 3.5.1. Notices that we either could use the fitting heuristic of Birgé and Massart (see subsections 3.4.1 and 3.5.1) in order to adapt the multiplicative term λ to the data. Here the goal of the method consists of approximate the correction term λ as :

$$\|Y - \hat{f}_k\|_n^2 \simeq -\frac{1}{2} \lambda \frac{k}{n} g_n(k) + \gamma$$

with $k_{\min} \leq k \leq k_{\max}$.

3.5.3 Second method in practice

This section aims to demonstrate the second method practice. We have applied the selection method with the same example as we have used for testing the first method (in section 3.4.2). We have simulated several ALT samples using the 3-arrhenius Weibull model draw in fig.3.1. for many sampling sizes : $n=50, 100, 200$ and 400 . We have implemented the data-driven correction of the penalty by using for each sampling size n the associated \hat{c} . We have compared the performances of PLSE, AIC, AICc, BIC and BICc. Let us notice that for each partition M_k candidate to be the best partition on k pieces, we have implemented a dynamic algorithm for PLSE and BICc as the criteria is associating with LSE and a given scale parameter (the variance term is correcting by the method. Otherwise AIC, AICc and BIC are based on MLE. So we have implemented an exhaustive research and in a way of simplicity, we have supposed that $k_{\max} = 10$ during the exhaustive research.

Table 3.5: The table summarizes the second method and it shows empirical distribution of the PLSE method with the classical ones.

ratio	PLSE	AIC	AICc	BIC	BICc
$n = 50 : k$	2 3 4 5	2 3 4 5	2 3 4 5	2 3 4 5	2 3 4 5
$\bar{p}_{\hat{k}}(k)$	0 54 6 4	0 6 5 4	0 6 5 4	17 32 2 1	0 61 16 8
$n = 100 : k$	2 3 4 5	2 3 4 5	2 3 4 5	2 3 4 5	2 3 4 5
$\bar{p}_{\hat{k}}(k)$	0 58 8 4	0 1 4 5	0 1 4 5	0 28 4 5	0 65 9 3
$n = 200 : k$	2 3 4 5	0 1 1 3	2 3 4 5	2 3 4 5	2 3 4 5
$\bar{p}_{\hat{k}}(k)$	0 73 6 6	0 1 1 3	0 1 1 3	0 36 1 1	0 79 4 4
$n = 400 : k$	2 3 4 5	2 3 4 5	2 3 4 5	2 3 4 5	2 3 4 5
$\bar{p}_{\hat{k}}(k)$	0 89 2 1	0 0 0 0	0 0 0 0	0 90 0 0	0 89 4 1

On the table 3.5 and the figures 3.7 and 3.8, we have got the empirical estimation of \hat{k} the estimating number of accelerating steps for each selection method. Since we are using the second method, the empirical distribution of k is concentrating on the true value of $k=3$. Next, it never underestimates the model and the rate of overestimation is decreasing with n . Anyways, the first method seems better than the second one if we are comparing through the empirical distribution of \hat{k} or referring with the estimation of the risk ratio. Otherwise the methods based on Akaike's penalty are strongly underestimating the penalty, see 3.7. And they are always overestimating the model.

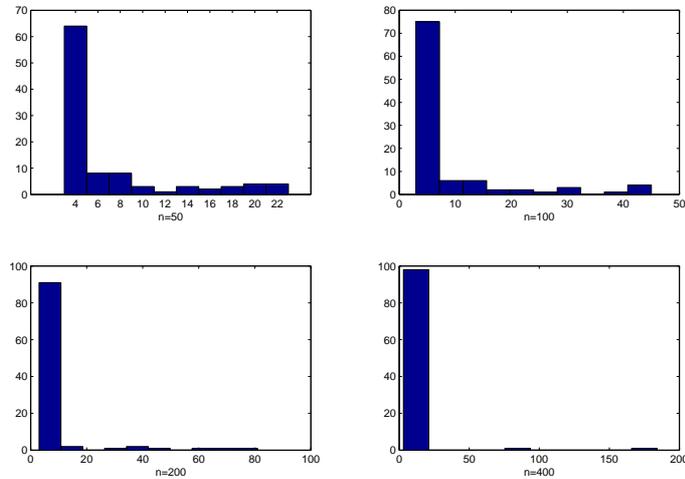


Figure 3.6: The histogram of \hat{k} based on PLSE criteria and for a 3 Arrhenius Weibull ALT.

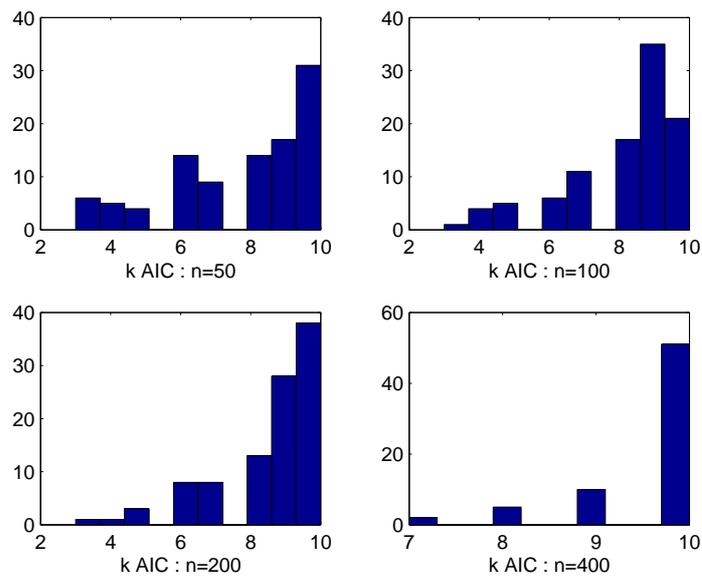


Figure 3.7: The histogramm of \hat{k} based on Akaike's selection method and for a 3 Arrhenius Weibull ALT.

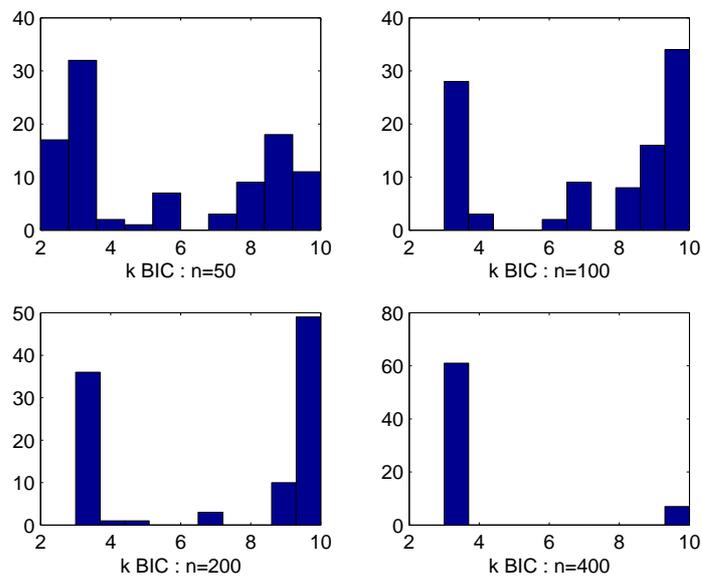


Figure 3.8: The histogramm of \hat{k} based on BIC criteria and for a 3 Arrhenius Weibull ALT.

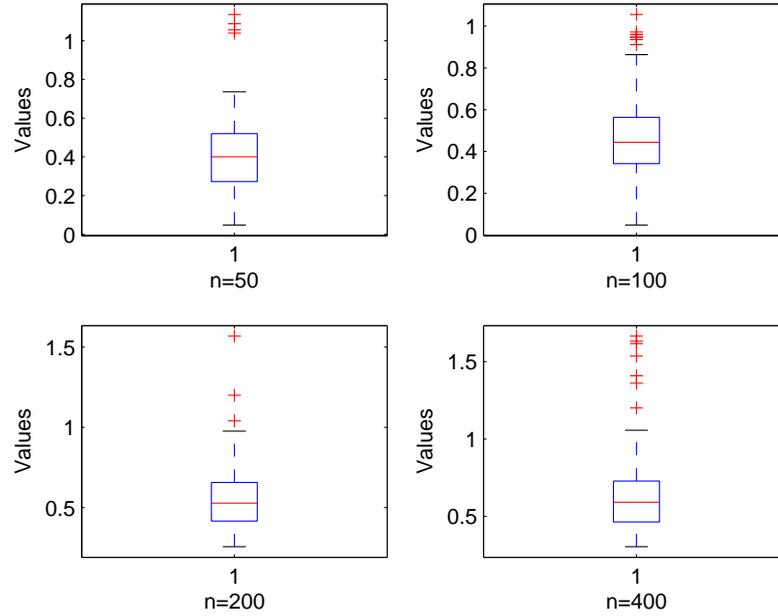


Figure 3.9: The boxplots of $\hat{\lambda}$ based on PLSE based on BIC penalty and for a 3 Arrhenius Weibull ALT.

Furthermore, the BIC criterion gives better results than AIC. But if n is not equal to 400, the empirical distribution are also concentrating around the extremes values of k . As we could see in figure 3.8, since $n \leq 200$, the true value $k=3$ and the dimension of the largest model is 10). And if BIC leads to overestimation, than we notice that for BICc, we find that $\alpha_{\min} > 0.5$. In figure 3.9, we see the box plots of correction for BIC's penalty. Here we clearly see that the the median of $\hat{\lambda}$ the data-driven correcting term is more than 0.5 when $n \geq 200$.

Table 3.6: The table summarizes the second method and it shows empirical risk ratio for the PLSE and the classical estimators.

	PLSE	AIC	AICc	BIC	BICc
$n = 50$	1.7013	2.4175	2.4175	3.5229	1.5195
$n = 100$	1.7598	2.1795	2.1795	1.8435	1.4362
$n = 200$	1.4845	1.9906	1.9877	1.6735	1.2877
$n = 400$	1.1605	1.7669	1.7669	1.0619	1.0940

And on the table 3.6, we have given an estimation of the risk ratio between the PLSE and the oracle of the grouped models (dimension per dimension). As in theoretical results, the selection method based on Sauvé's theorem gives the better results in terms of risk ratio. Let us notice that the calibration based on the calibration of c gives good results : the risk ratio keeps to 1 and the ratio are closer to 1 as n the sampling size is

increasing. Also, the risk ratio of the corrected criteria based on penalty of BIC has the same magnitude as the PLSE. And next, the risk ratio for criterion of AIC, AICc and BIC are clearly larger than the new PLSE, except for BIC and when $n=400$. Lastly, the fact that the estimation based on AIC or BIC are far from the oracle estimation clearly explains why the PLSE of the table 3.5 detects the ruptures on the data.

3.5.4 Class-H motor insulation data

Now we have considered real data give by Nelson [27], chapter 7. The test units of class-H data are motor insulation. The stress covariable is the temperature (with four temperatures 190, 220, 240 and 260°C). We assume the acceleration relationship is Arrhenius one. And we consider that the lifetimes are Weibull data. As it was made on class-H data, the experiment keeps on running while the first fail occurs. So we observe three failures roots causes: turn, phase and ground. Nelson proposes to analyse each failure mode independently and he uses a series-system model. He has also remarked that the turn fail occurs on lower stress levels and the three fails are on competition for higher stress levels. For this application, we only have used the first fail and we propose to detect the changes on the acceleration between lower and higher stress levels by our selection method. Let us notice that such an analysis may not replace some competing risk analysis, here our aim is only to underline the misleading on a classical ALT.

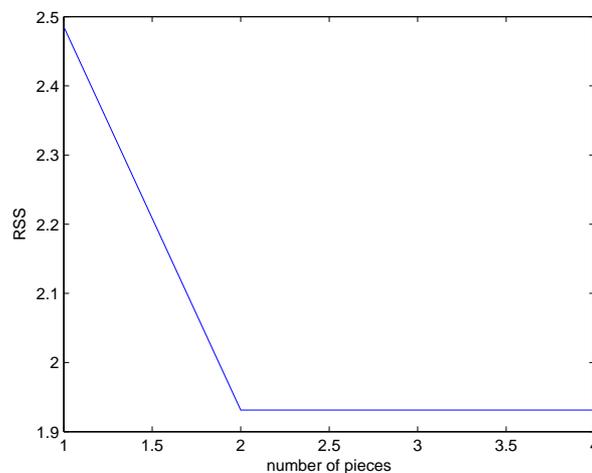


Figure 3.10: The RSS of the piecewise regression models \hat{M}_1 , \hat{M}_2 , \hat{M}_3 and \hat{M}_4 .

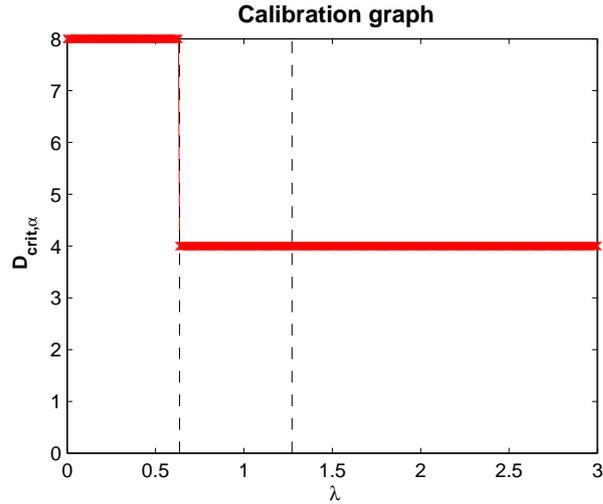


Figure 3.11: The calibration plot represents the selecting dimension when the correcting term α varies.

The sampling size is 40, we assume that $c = 1.218$, thanks to the estimation of section 3.5.1 based on full noisy data. And we have also considered that the plausible ruptures on the acceleration may occur on temperatures 200, 230 and 250°C. We have run on the dynamic algorithm (see appendix A) and the heuristic trend (see 3.5.2). The dynamic algorithm gives the following stress sets :

$$\begin{aligned}\hat{M}_1 &= \{ [0, +\infty[\} \\ \hat{M}_2 &= \{ [0, 230[, [230, +\infty[\} \\ \hat{M}_3 &= \{ [0, 200[, [200, 230[, [230, \infty[\} \\ \hat{M}_4 &= \{ [0, 200[, [200, 230[, [230, 250[, [250, \infty[\}\end{aligned}$$

In figure 3.10, we have traced the residual sum of square of the models candidates to the estimation (indexing by the number of ruptures on the acceleration relationship). Next, the figure 3.11 represents the selecting dimensions versus α the correction term using during the heuristic trend. We observe that the minimal dimensional gap occurs when $\alpha_{\min} = 0.636$, so $\alpha_{\text{opt}} = 1.272$ and the selected model is \hat{M}_2 . Such a choice seems logical, because the model regroups the stress 190 and 220°C where the turn fail occurs. And, on the other hand, it regroups the stress 240 and 260°C where we notice competition between the three failure modes. In this example, the selection method have detected the changes between lower a higher stress levels. At the end, we have traced the Class-H data and the PLSE estimation on figure 3.12, and we have also traced the residual plot on figure 3.13.

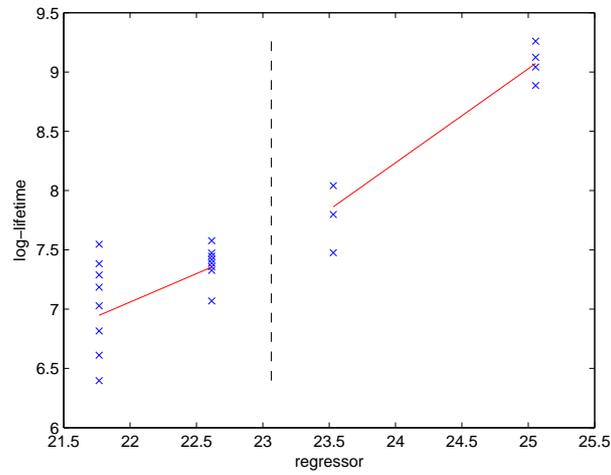


Figure 3.12: The blue crosses represent the observations of class-H insulation data and the red line shows the PLSE.

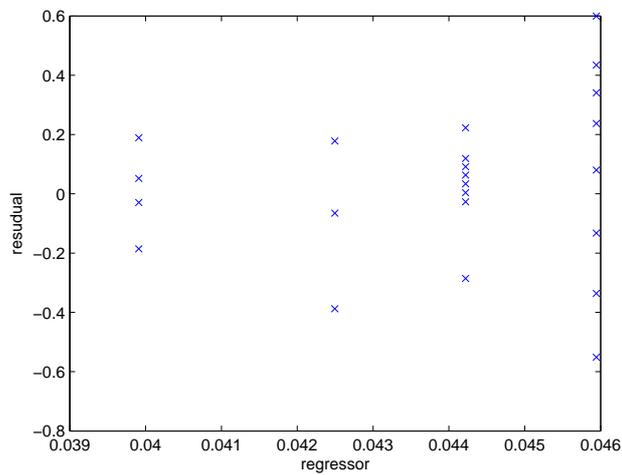


Figure 3.13: The residual plot for PLSE.

Conclusion et perspectives

Notre travail s'est axé autour des tests accélérés sous environnements constants. Plus précisément, nous avons traité de l'estimation des lois d'accélération à partir d'essais accélérés réalisés sur des composants issus de hautes technologies comme les semi-conducteurs.

Dans ce cadre, nous avons utilisé les outils de sélection de modèles proposés par Birgé et Massart (cas Gaussien) et les récents travaux de Sauvé (cas Weibull). Nous avons ainsi étudié deux cas typiques de fiabilité :

1. Lorsque le mode de défaillance est dû à l'usure des matériaux, la loi de durée de vie est supposée Lognormale. Dans ce cas, nous avons appliqué les résultats théoriques de sélection de modèles par la méthode des moindres carrés pénalisés de Birgé et Massart. Nous avons proposé des méthodes de calibration du critère pénalisé adaptées à l'estimation des lois d'accélération et nous avons appliqué ces outils de sélection sur des données simulées et des données réelles de référence [27].
2. Lorsque les modes de défaillance surviennent à la suite d'un choc, la loi de durée de vie est la loi de Weibull. Nous avons donc utilisé les résultats théoriques de sélection de modèles par la méthode des moindres carrés pénalisés de Sauvé. Comme pour le cas Lognormal, nous avons mis en place des procédures de calibration adaptés aux tests accélérés et nous avons appliqué ces méthodes sur des données simulées et des données réelles de référence.

Ces deux outils de sélection de modèles permettent à l'ingénieur de réaliser ses études de fiabilité sans *a priori* sur les modèles d'accélération. L'ingénieur fiabiliste dispose d'un certain nombre d'outils logiciels afin d'analyser les essais de fiabilité, comme *SAS*, *ALTA*, *Weibull++*, ... Il peut se servir de ces outils afin de sélectionner un modèle à partir d'un critère AIC. Comme nous l'avons vu dans nos études, le critère AIC a tendance à surestimer les modèles et donc à sélectionner des modèles bien trop complexes. Nous proposons dans notre travail d'utiliser un outil performant d'analyse des essais accélérés adapté à l'objectif d'amélioration des procédés ou des changements technologiques. Au même titre qu'AIC, la mise en oeuvre de nos outils de sélection ne nécessitent pas un trop lourd travail de programmation, pour notre part nous avons créé une série de programmes sous *MatLab* afin de réaliser l'ensemble des simulations et des applications.

Ensuite dans les domaines de hautes technologies, il y a de constants changements technologiques liés aux innovations concernant les matériaux, l'architecture ou les procédés industriels. Ces innovations influencent fortement la loi d'accélération. Dès lors, il paraît déraisonnable d'utiliser le ReX ou des plans d'expérience issues de standards de fiabilité. Dans ce cas, nos outils de sélection permettent de valider les modèles à la suite d'un tel changement technologique et d'éviter les erreurs d'*a priori* liés aux modèles.

Par ailleurs, ces méthodes de sélection permettent de détecter les changements de comportements moyens hors des limites de tolérance. Dans ce genre de cas, on surcharge volontairement les stress utilisés lors de l'essai. Ces tests accélérés sont des tests HALT, pour High Accelerating Life Test. Ils sont souvent réalisés dans les articles de physique de la défaillance afin d'explorer les limites d'utilisation des composants. Dans ces études, l'estimation des caractères de fiabilités dans des conditions environnementales extrêmes permet de déterminer quelles sont les innovations de procédés, de matériaux ou d'architectures les plus performantes. Par ailleurs, lorsque l'on travaille sur les défaillances de chocs, les niveaux de stress peuvent être très supérieurs aux niveaux de tolérance. Dès lors, il paraît raisonnable d'estimer les nouvelles tolérances associées aux conditions de stress. Grâce aux méthodes de sélection, nous disposons d'un outil de prévision adapté à ces deux cas de figure.

En outre, les méthodes de sélection de modèle prennent en compte un large spectre de lois de probabilité. Ici, nous avons considéré les lois de durées de vie de Weibull et Lognormale, cependant les résultats de sélection de Sauvé offrent un cadre théorique non-Gaussien dans un large spectre de lois de probabilité. Plus précisément, le théorème émet l'hypothèse que la transformée de Laplace de la loi de probabilité est bornée au voisinage de zéro.

Suite à ce travail, on peut envisager un certain nombre de perspectives. Tout d'abord, notre approche utilise l'estimateur des moindres carrés pénalisé. Un certain nombre d'études ont utilisé l'estimateur du maximum de vraisemblance pénalisé en se basant sur l'approche non-asymptotique de Birgé et Massart (par exemple Castellan [11]) et un critère construire à partir d'une décomposition du risque Kullback ou de la distance d'Hellinger. Ce type d'approche est important dans le cas de données censurées, lors de tests accélérés il arrive de rencontrer des censures de type I ou même des censures par intervalles.

Et d'autre part, tout au long notre étude, le critère de sélection compare les modèles à partir du risque des estimateurs des moindres carrés des modèles candidats. Il serait intéressant d'explorer une approche basée sur l'estimation du taux de défaillance. Il faudrait alors reprendre l'idée directrice des travaux de Birgé et Massart afin d'établir une inégalité oracle qui permettrait de contrôler la distance entre le taux de défaillance *réel* et le taux de défaillance de l'estimateur pénalisé.

Appendix A

Recherche des modèles candidats

A.1 Partition dynamique

Soit (Y, \mathbf{x}) un n -échantillon de régression tel que :

$$Y_i = f(\mathbf{x}_i) + \sigma \varepsilon_i$$

où f est une fonction de régression (constante, affine ou polynomiale) définie par morceaux sur l'ensemble des covariables \mathcal{X} , σ^2 est le paramètre de variance et $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)$ est le vecteur aléatoire des erreurs.

En vue de sélectionner le meilleur modèle de régression défini par morceaux, nous considérerons une partition fine, notée M_0 . Et nous comparerons les modèles définis par les partitions M construites à partir des éléments de M_0 . Nous noterons \mathcal{M}_n l'ensemble des partitions M construites à partir des éléments de M_0 . Comme nous l'avons vu dans les parties 2 et 3, le critère de sélection par pénalisation d'une partition M est une fonction du nombre k de morceaux disjoints de la partition M . Ainsi, à k fixé, notons M_k comme étant le meilleur modèle à k morceaux. Alors le critère pénalisé sélectionnera les modèles définis par les partitions $M_1, M_2, \dots, M_{N_n} = M_0$.

Lorsque la partition M_0 est grande, la recherche des partitions $M_1, M_2, \dots, M_{N_n-1}$ est très longue (comme on réalise une recherche exhaustive). On utilise un indice statistique afin de sélectionner la meilleure partition candidate pour une dimension donnée. Comme nous l'avons vu précédemment, nous définissons le meilleur candidat comme étant celui qui minimise la somme du carré des résidus (SCR). Et compte tenu du fait que la fonction de régression est définie par morceaux selon d'axe des régresseurs, il apparaît possible de diminuer la complexité algorithmique par une programmation dynamique. Lorsque le régresseur est univarié, soient r_0, r_1, \dots, r_{N_n} les $(N_n + 1)$ points de rupture définissant M_0 , avec les conventions $r_0 = \inf_{\{i=1, \dots, n\}}(\mathbf{x}_i)$ et $r_{N_n} = +\infty$, de sorte que les éléments de M soient les intervalles semi-ouverts $[r_{j-1}, r_j[$ pour $j = 1, \dots, N_n$. Définissons $\Delta(i, j)$ la SCR du modèles de régression défini sur l'intervalle $[r_i, r_j[$. Et définissons $I_k(j)$ la valeur minimale de SCR du modèles de

régression à j morceaux de l'intervalle $[r_0, r_j[$, alors :

$$I_k(j) = \inf_{i_0=0 < i_1 < \dots < i_k=N_n} \sum_{l=1}^k \Delta(i_{l-1}, i_l).$$

Comme les régressions sont réalisées partition par partition, il vient la formule récursive suivante :

$$I_k(j) = \inf_{\{1 \leq j' \leq (j-1)\}} \left[I_{k-1}(j') + \Delta(j', j) \right]. \quad (\text{A.1})$$

Grâce à propriété récursive (A.1), il apparaît possible d'optimiser la recherche de partition par un algorithme dynamique de complexité en $o(n^2)$:

ALGORITHME DYNAMIQUE:

Consistué d'une étape d'initialisation et d'étapes récursives.

(E₀) Calculer $\Delta(i, j)$ pour $0 \leq i < j \leq N_n$;

(E₁) Calculer $I_1(j) = \Delta(0, j)$ pour $1 \leq j \leq N_n$
Stocker les indices $i_0^{(1,j)} = 0$ et $i_1^{(1,j)} = j$;

⋮

(E_k) Calculer $I_k(j)$ pour $0 \leq k, j \leq N_n$ en utilisant (A.1):
Déterminer $j_k = \arg \inf_{\{1 \leq j' \leq (j-1)\}} [I_{k-1}(j') + \Delta(j', j)]$
Stocker les indices $i_0^{(k,j)} = i_0^{(k-1,j_k)}$, $i_1^{(k,j)} = i_1^{(k-1,j_k)}$, ...,
 $i_{k-1}^{(k,j)} = i_{k-1}^{(k-1,j_k)} = j_k$ et $i_k^{(k,j)} = j$;

⋮

Au terme de la boucle, on conserve les indices $i_0^{(k,N_n)} = 0 < i_1^{(k,N_n)} < \dots < i_{k-1}^{(k,N_n)} < i_k^{(k,N_n)} = N_n$ des ruptures minimisant la SCR en $I_k(N_n)$ pour $k = 1, \dots, N_n$.

A.2 Partition CART

Une méthode classique de sélection de modèles par morceaux est la régression CART (cf. Freidman et al.). La méthode CART se décompose en deux étapes :

1. Détermination des partitions emboîtées ;
2. Sélection de la partition.

La première étape permet de trouver des partitions emboîtées par une méthode récursive. À chaque étape, on effectue une partition dichotomique d'un sous-ensemble de la partition l'ensemble. La partition à *splitter* et le point de rupture sont choisis en fonction d'indices statistiques, à savoir : en minimisant la SCR. Et à terme, on obtient une suite de partitions emboîtées que l'on représente souvent en forme d'arbre. C'est ainsi que l'on présente souvent cette étape comme la détermination d'un arbre maximum, noté T_0 . La seconde étape consiste à sélectionner une partition par validation croisée, en référence au vocabulaire de la représentation par arbres, on parle d'une étape d'élagage de l'arbre maximal T_0 . Dans le cadre de notre étude, nous utiliserons principalement la première étape CART comme le sujet de l'étude est d'appliquer un critère sélection par pénalisation dimensionnelle.

Lorsque les données de régression (Y, \mathbf{x}) sont telles que l'ensemble des covariables est de dimension $p \geq 1$, posons $\mathbf{x} = (\mathbf{x}_1, \dots, \mathbf{x}_n)$ et $\mathbf{x}_i = (x_{i,1}, \dots, x_{i,p})$. Lorsque $j = 1$, on utilise l'algorithme dynamique. Malgré l'algorithme dynamique, lorsque $p \geq 2$, il devient long et complexe de déterminer la collection exhaustive des partitions de la régression par morceaux. On paraît raisonnable de limiter le nombre de modèles candidats en utilisant une méthode de recherche optimisée de partition telle que CART. L'algorithme CART permet de construire des partitions emboîtées par une méthode récursive en utilisant un critère statistique.

Le principe récursif de CART est le suivant : supposons que l'on dispose de k régions R_1, \dots, R_k partitionnant l'ensemble des covariables, pour une région R_l donnée et un axe $1 \leq j \leq p$, on définit deux régions : $R_1(j, r) := \{x_i : x_{i,j} < r\}$ et $R_2 := \{x_i : x_{i,j} \geq r\}$. On sélectionne la partition dichotomique en $R_1(j, r)$ et $R_2(j, r)$ optimale pour chacune des k régions R_1, \dots, R_k . On obtient donc une suite d'indices $(j_1, r_1), \dots, (j_k, r_k)$ définissant la partition optimale de chaque région. Et on choisit l'indice $(j_{\hat{k}}, r_{\hat{k}})$ de telle sorte que la SCR du modèle de régression associé à la partition (en $(k+1)$ morceaux) $R_1, \dots, R_1(j_{\hat{k}}, r_{\hat{k}}), R_2(j_{\hat{k}}, r_{\hat{k}}), \dots, R_k$ soit minimale.

En résumé, voici les étapes principales de l'algorithme de construction de l'arbre maximale CART :

ALGORITHME CART:

Construction de la partition maximale par une étape de Split et puis par étapes récursives en utilisant le Split.

(E₁) Pour tout $j = 1, \dots, p$: définir des régions de split :

$$R_1(j, r) := \{x_i : x_{i,j} < r\} \text{ et } R_2 := \{x_i : x_{i,j} \geq r\}.$$

Déterminer l'axe j_1 et la rupture r_1 minimisant $I_1(1)$ la SCR du modèle de partition en deux régions $R_1 := R_1(j, r)$ et $R_2 := R_2(j, r)$;

⋮

(E_k) Supposons $(k - 1)$ régions R_1, \dots, R_{k-1} :

- Pour tout $l = 1, \dots, (k - 1)$: définir l'axe $j_{k,l}$ et la rupture $r_{k,l}$ minimisant $I_k(l)$ la SCR du modèle de partition défini en utilisant les régions R_1, \dots, R_{k-1} et en splittant la région R_l en $R_1(j_{k,l}, r_{k,l})$ et $R_2(j_{k,l}, r_{k,l})$,
- Définir la région \hat{l} à splitter en minimisant $I_k(l)$,
- Construction de la partition en k régions R_1, \dots, R_k : Retenir toutes les régions R_1, \dots, R_{k-1} sauf celle indexée par \hat{l} et on rajoute à la liste les régions $R_1(j_{k,\hat{l}}, r_{k,\hat{l}})$ et $R_2(j_{k,\hat{l}}, r_{k,\hat{l}})$;

⋮

La boucle s'achève lorsque l'opération de Split mène à un modèle de régression par morceaux trivial.

List of Figures

1.1	Schéma du transistor MOS (pour Metal Oxide Semiconductor) en fonction sur les schémas ; en a/ le courant est stoppé comme la tension sur la grille est nulle et en b/ on applique du courant à la grille et le transistor laisse passer le courant.	15
1.2	Schéma de coupe d'un circuit intégré moderne à plusieurs niveaux de connexions métalliques.	16
1.3	a/ Puce microélectronique intégrée sur une plaque de silicium; et b/ Zoom avant sur la surface de la puce.	17
1.4	Schéma d'organisation de la ligne de production.	18
1.5	Schéma simplifié des processus d'intégration nécessaires à la fabrication d'un transistor.	21
1.6	En a) évolution de la résistance ; et en b) évaluation des performances d'un transistor durant une période de test.	24
1.7	Deux configurations dans l'évolution de la résistance durant un test: a) un lent accroissement ; b) un pic de résistance.	27
1.8	Schéma de l'électromigration d'une interconnexion.	28
1.9	Schéma de stressmigration d'une interconnexion.	29
1.10	Relation entre la tension source-grille et la variation de l'intensité drain/source ; la courbe violette représente le transistor en fonctionnement usuel et la courbe bleu clair montre le transistor en fonctionnement dégradé. . .	31
1.11	Schéma du claquage diélectrique d'un transistor.	32
1.12	Représentation de la fonction de densité de probabilité ϕ d'une loi Log-normale pour différentes valeurs du paramètre de forme $\sigma = 1,5 ; 0,5 ; 0,25$ et $\mu=2$	35
1.13	Représentation de la fonction de densité de probabilité ψ d'une loi Weibull pour différentes valeurs du paramètre de forme $\beta = 0,5 ; 1 ; 1,5 ; 3$ et $\eta=2$. . .	38
1.14	Principe du test ALT.	39
1.15	Exemple d'essai accéléré Arrhénius Lognormal estimé par les moindres carrés: en a) on retrouve le nuage de points et la droite de régression ; et en b) le graphique des résidus.	49

1.16	Exemple d'essai accéléré puissance-Weibull estimé par les moindres carrés : en a) on retrouve le nuage de points et la droite de régression ; et en b) le graphique des résidus.	50
1.17	Exemple d'essai accéléré puissance-Weibull estimé par le maximum de vraisemblance, résultats d'un bootstrap paramétrique de l'EMV sur 300 répétitions : en a) on retrouve le boxplot des paramètres $\hat{a}_{0,MV}$ et $\hat{a}_{1,MV}$; et en b) le boxplot de $\hat{\sigma}_{MV}$	51
1.18	Distribution de données accélérées lorsqu'il y a deux plages d'accélération masquées, on observe un « plateau » représentant le changement de mode.	51
1.19	Exemple d'un modèle d'accélération à plusieurs plages d'accélération.	53
2.1	The empirical estimation of the risk ratio for piecewise Arrhenius accelerating Gaussian model : we have plot out $\bar{M}_n(c_1, c_2)$ when $c_2 = 1, 2, 3, 3.4, 4$ and 5 in function of c_1	67
2.2	Contrast of the risk ratio $\bar{M}_n(c_1, c_2)$ for all $n \in \mathcal{N}$ for Gaussian model.	68
2.3	Maximum values of contrast of the risk ratio $\{\bar{M}_n(c_1, c_2) - \inf_{(\alpha, \beta) > 0} (M_n(\alpha, \beta))\}$ for Gaussian model.	69
2.4	Boxplot for c for various series of sampling sizes, in 1 $5 \leq k \leq n/4$, in serie 2 $5 \leq k \leq n/3$, ... and n varies from 50 to 400.	73
2.5	Boxplot for λ for various series of sampling sizes, in 1 $5 \leq k \leq n/4$, in serie 2 $5 \leq k \leq n/3$, ... and n varies from 50 to 400.	73
2.6	Boxplot for c for various series of sampling sizes, in 1 $n = 50$, in 2 $n = 100$, in 3 $n = 200$ and in 4 $n = 400$ and k_{max} varies form $n/5$ to $n/2$	75
2.7	Representing the variations of the penalized estimation when the correction term α varies.	77
2.8	Box plot of correction since the penalty function of BIC for all sampling sizes.	79
2.9	The histogramm of \hat{k} based on PLSE1 and for a 3 Arrhenius Lognormal ALT.	81
2.10	The histogramm of \hat{k} based on PLSE2 and for a 3 Arrhenius Lognormal ALT.	81
2.11	The histogramm of \hat{k} based on Akaike's selection method and for a 3 Arrhenius Lognormal ALT.	82
2.12	The histogramm of \hat{k} based on BIC criteria and for a 3 Arrhenius Lognormal ALT.	82
2.13	The histogramm of \hat{k} based on BICc criteria and for a 3 Arrhenius Lognormal ALT.	83
2.14	The RSS of the piecewise regression models $\hat{M}_1, \hat{M}_2, \hat{M}_3$ and \hat{M}_4	84
2.15	The calibration plot represents the selecting dimension when the correcting term α varies.	84

2.16	The blue crosses represent the observations of class-B insulation data and the red line shows the PLSE.	85
2.17	The residual plot for PLSE.	85
2.18	The fit of the penalty function as minus half of the euclidean norm of the LSE for $\hat{M}_3, \hat{M}_4 \dots \hat{M}_{10}$	87
2.19	The calibration plot represents the selecting dimension when the correcting term α varies.	87
2.20	The blue crosses represent the observations of voltage drop data and the red line shows the PLSE.	88
2.21	The blue crosses represent the observations of voltage drop data and the red line shows the piecewise cubic estimation with two ruptures on the times 6.5 and 13.	89
2.22	The residual plot for PLSE.	89
3.1	The plot is drawing (red line) the 3-piecewise acceleration function using for the simulations, and we have also drawn (blue crosses) the log-lifetimes of a simulating sample using Gumbel's distribution.	102
3.2	Boxplot for c when the lifetime distribution is Weibull and various series of sampling sizes, in 1 $5 \leq k \leq n/4$, in 2 $5 \leq k \leq n/3$, ... and n varies from 50 to 400.	102
3.3	Boxplot for λ when the lifetime distribution is Weibull and various series of sampling sizes, in 1 $5 \leq k \leq n/4$, in 2 $5 \leq k \leq n/3$, ... and n varies from 50 to 400.	103
3.4	Boxplot for c when the lifetime distribution is Weibull and various series of sampling sizes.	107
3.5	The heuristic trend in practice for 5 piecewise Arrhenius Weibull problem.	109
3.6	The histogramm of \hat{k} based on PLSE criteria and for a 3 Arrhenius Weibull ALT.	110
3.7	The histogramm of \hat{k} based on Akaike's selection method and for a 3 Arrhenius Weibull ALT.	111
3.8	The histogramm of \hat{k} based on BIC criteria and for a 3 Arrhenius Weibull ALT.	111
3.9	The boxplots of $\hat{\lambda}$ based on PLSE based on BIC penalty and for a 3 Arrhenius Weibull ALT.	112
3.10	The RSS of the piecewise regression models $\hat{M}_1, \hat{M}_2, \hat{M}_3$ and \hat{M}_4	113
3.11	The calibration plot represents the selecting dimension when the correcting term α varies.	114
3.12	The blue crosses represent the observations of class-H insulation data and the red line shows the PLSE.	115
3.13	The residual plot for PLSE.	115

List of Tables

1.1	Liste de quelques semi-conducteurs.	14
1.2	Évolution des semi-conducteurs par une liste de quelques processeurs INTEL.	18
1.3	Tableau dressant une liste non exhaustive de défauts en décrivant en détail la cause et le niveau d'intégration.	25
1.4	Modèles accélérés classiques.	45
2.1	This table gives us the value of $\bar{M}_n(c_1, c_2)$ when n varies and for some pairs (c_1, c_2) candidates to the calibration.	68
2.2	The table represents the median and mean estimations of c^* for Lognormal ALT with various sampling sizes.	72
2.3	The table represents $\bar{p}_{\hat{k}}(k) = \frac{1}{100} \sum_{j=1}^{100} 1_{(\hat{k}(y^{(j)}, \mathbf{x}^{(j)})=k)}$ the empirical distribution of k using the PLSE for Lognormal ALT with various sampling sizes.	72
2.4	The table represents the median and mean estimations of c^* based on the pure noisy method for Lognormal ALT with various sampling sizes.	75
2.5	Table of the empirical distribution of \hat{k} the penalized estimator of the number of ruptures on the data and for various selection procedures: AIC, BIC, AICc and the adaptation penalty with the three data-driven methods : PLSE1, PLSE2 and BICc.	79
2.6	Table of the empirical risk ratio between the estimation risk and the oracle risk for all tested procedures: AIC, BIC, AICc and our penalty with the three data-driven methods : PLSE1, PLSE2 and BICc.	80
3.1	Presents a list of non exhaustive life-stress relationship.	92
3.2	The table represents the median estimations of c^* and λ^* and the empirical risk ratio for Weibull ALT with various sampling sizes.	103
3.3	The table represents the empirical distribution of k using the PLSE for Weibull ALT with various sampling sizes.	104
3.4	The table represents the median estimations of c^* based on the pure noisy method for Weibull ALT with various sampling sizes.	106

3.5	The table summarizes the second method and it shows empirical distribution of the PLSE method with the classical ones.	110
3.6	The table summarizes the second method and it shows empirical risk ratio for the PLSE and the classical estimators.	112

Index

- algorithme
 - CART, 60, 108
 - recherche dynamique, 61, 89, 105–106
- ALT, 54, 82
 - essais accélérés, 33, 39
 - hypothèse, 39–41, 54, 82
 - multiples accélérations, 50, 55, 57, 84, 89, 91–96
 - relation d'accélération, 54, 56, 82
- breakdown, 48, 82
- calibration, 62
 - Birgé et Rozenholc, 63–64, 95
 - correction de pénalité, 74, 98
 - méthode Birgé-Massart, 90–94
 - pur bruit, 72, 95–98
- critère pénalisé, 57–59
 - Akaike, 58, 86
 - Akaike corrigé, 58
 - correction de pénalité, 98
 - fonction de pénalité, 60, 61, 88, 89, 94
 - général, 86
 - Mallows, 86
 - Schwarz, 58, 86
- distribution de probabilité
 - Gumbel, 37, 85
 - Lognormale, 34–36, 55
 - lognormale, 36
 - Weibull, 36–38, 83, 96
- electromigration, 26–28, 54, 82
- estimateur
 - maximum de vraisemblance, 44–45, 84
 - moindres carrés, 43–44, 57, 84, 85
 - moindres carrés pénalisé, 58, 84, 87, 89, 91
 - ridge, 44
- heuristique
 - Birgé et Rozenholc, 62
 - Birgé-Massart, 62, 68, 90–91, 96, 98–99
 - Mallows, 68, 90
- mode de défaillance
 - migration métallique, 26, 28, 29
 - claquage diélectrique, 31–32
- oracle, 86
 - inégalité oracle, 88, 95
- physique de la défaillance, 22
- retour d'expérience, 32–33
- risques compétitifs, 50
- stressmigration, 26, 29–30, 54, 82
- théorème
 - sélection de modèles, 87
 - sélection de modèles Gaussiens, 59

Bibliography

- [1] Akaike H. (1973), *Information theory and an extension of the maximum likelihood principle*, Proceedings of 2nd international symposium on information, pp. 267-281, Budapest.
- [2] Bacha M., Celeux G., Idée E., Lannoy A., Vasseur D. (1998), *Estimation de modèles de durées de vie fortement censurées*, Coll. de la direction des études et recherches d'Electricité De France, éd. Eyrolles.
- [3] Bagdonavicius V., Gerville-Réache L., Nikouline V., Nikulin M. (2000), *Expériences accélérées: analyse statistique du modèle standard de vie accélérée*, Rev. Stat. Appl., XLVIII (3), pp. 5-38.
- [4] Baraud Y., Comte F. and Viennet G. (2001), *Adaptation estimation in autoregression or-mixing regression via model selection*, Ann. stat., vol. 29, 3, pp. 839-879.
- [5] Barron A., Birgé L., Massart P. (1999), *Risk bounds for model selection via penalization*, Prob. Theory & Related Fields, 113, pp. 301-413.
- [6] Birgé L., Massart P. (2001), *Gaussian Model Selection*, J. of the Europ. Math. Soc., 3, pp. 203-268.
- [7] Birgé L., Massart P. (2001), *A Generalized C_p Criterion for Gaussian Model Selection*, prépubli. 647, universités Paris-VI & Paris-VII.
- [8] Birgé L., Rozenholc Y. (1999), *How many bins should be put in a regular histogram*, Tech. Report, Publication Université Paris-VI.
- [9] Black J. (1969), *Electromigration a brief survey and some recent results*, IEEE Transactions on Electronic Devices, vol. 16, 4.
- [10] Bon J-L. (1995), *Fiabilité des Systèmes - Méthodes mathématiques*, coll. Techn. Stoch., Masson.
- [11] Castellan G. (2001), *Estimation d'une densité par sélection de modèles exponentiels*, C. R. Acad. Sci. Paris, t. 332, Série I, pp. 465-468.
- [12] Coles S., (2001), *An introduction to statistical modeling of Extrem Values*, Springer series in Stat., Springer-Verlag London.

-
- [13] Crowder M.J., Kider A.C., Smith R.L., Sweeting T.J. (1991), *Statistical Analysis Reliability Data*, Chapman and Hall.
- [14] Hastie, T., Tibshirani, R., Friedman, J. (2001), *The Elements of Statistical Learning* Springer-Verlag.
- [15] Hurvich, M.C., Tsai, C.L. (1989), *Regression and time series models selection in small samples*, *Biometrika*, 76, pp. 297-307.
- [16] Jensen F., Petersen N.E. (1982), *Burn-In : An Engineering Approach to the Design and Analysis of Burn-in Procedures*, Wiley - Interscience Publication, John Wiley and Sons.
- [17] Lavarde M., Pamphile P. (2005), *Models Selection in Accelerated Life Tests*, Qualita 2005 - 6ème Congrès International Pluridisciplinaire Qualité et Sécurité de Fonctionnement, vol. 2, pp. 473.
- [18] Lavarde M., Pamphile P. (2006), *Performances of Penalized Model Selection Practice to Accelerating Life Test*, LambdaMu 15 - Maîtrise des Risques et Sécurité de Fonctionnement, Lille - Octobre 2006 - Session 3C: modélisation méthodologie.
- [19] Lebarbier E. (2002), *Quelques approches pour la détection de ruptures à horizon fini*, PhD thesis, Université Paris XI.
- [20] Lilen H. (2004), *Une brève histoire de l'électronique*, ed. Vuibert.
- [21] Massart, P. (2005), *Concentration inequalities and model selection*, Lecture Notes in Mathematics series, Springer.
- [22] Mallows, C. (1974), *Some comments on Cp*, *Technometrics*, 15, pp. 661-675.
- [23] Mann N., Schafer R. and Singpurwalla N. (1974), *Methods for Statistical Analysis of Reliability and Life Data*, John Wiley & Sons, pp. 419-461.
- [24] McPherson J., Dunn C., (1987), *A Model for Stress-Induced Metal Notching and Voiding in Very-Large-Scale-Integrated Al-Si (1%) Metallization*, *Journal of Vacuum Science and Technologie*, vol. B5, pp. 1321-1325.
- [25] Meeker P.Q., Escobar L.A. (1998), *Statistical Models for Reliability Data*, John Wiley.
- [26] Meeker P.Q., Escobar L.A. (1998), *Tutorial : Pitfalls of Accelerated Testing*, *IEEE Transactions on Reliability*, vol. 47, 2, pp114-118.
- [27] Nelson W. (1990), *Accelerated Testing: Statistical Models, Test Plans and Data Analysis*, Wiley series in Probability & Mathematical Statistics.
- [28] Nikulin M., Gerville-Réache L., Couallier V. (2007), *Statistiques des essais accélérés*, coll. meth. stoc. appli., éd. Hermes Science Publications.
- [29] Ogawa E., Lee KD., Blaschke V., Ho P. (2002), *Electromigration Reliability Issues in Dual-Damascene Cu Interconnections*, *IEEE Reliability*, 51, 4, pp. 403-419.

-
- [30] Rao, C.R., Toutenburg, H. (1995), *Linear Models : Least Squares and Alternatives*, Springer Series in Statistics, Springer-Verlag N.Y.
- [31] Sauvé M. (2006), *Piecewise polynomial estimation of a regression function*, pré-publication 2006-39 de l'Université Paris Sud, soumis à IEEE Transactions on Information Theory.
- [32] Sauvé M. (2006), *Sélection de modèles de régression non gaussienne, Application à la sélection de variables et aux tests de survie accélérés*, PhD thesis, Université Paris XI.
- [33] Schwartz, G. (1978), *Estimating the dimension of a model*, Ann. Stat., 6, pp. 461-464.
- [34] Schwarz, J.A. (1987), *Effect of temperature on the variance of the log-normal distribution of failure times due to electromigration damage*, J. Appl. Phys., 61, 2, pp.801-803.
- [35] Van Der Waart, A. (1998), *Asymptotic Statistics*, Cambridge University Press.
- [36] Wolf S. (1990), *Silicom Processing for VLSI Area - vol2 : Process Integration*, Lattice Press, California.

Résumé : Cette thèse traite de l'exploitation de données accélérées et de la sélection de modèles de régression dans un domaine de hautes technologies : les semi-conducteurs. Les données recueillies à la suite d'un test accéléré sont des données de régression. L'objectif du test est d'ajuster le comportement moyen du logarithme des durées de vie à l'aide d'une fonction f , dite fonction d'accélération. Cependant les données accélérées ont parfois des comportements complexes. Afin d'adapter la modélisation à ces comportements atypiques, nous avons cherché à détecter les changements de comportement de la fonction d'accélération. Nous proposons d'utiliser une collection de modèles de régressions définis par morceaux, pour chaque modèle candidat à l'estimation nous calculons l'estimateur des moindres carrés. Et nous sélectionnons le modèle final \tilde{f} à l'aide d'un critère des moindres carrés pénalisés. L'estimateur pénalisé est une approximation optimale du modèle réel au sens où le risque de l'estimateur pénalisé est comparable au risque minimum parmi l'ensemble des modèles candidats. De plus, nous disposons d'une borne de risque non asymptotique. Et nous avons cherché à limiter les hypothèses de modélisation afin de prendre en compte un grand nombre de cas pratiques : nous avons envisager le cas d'usure (loi de durée de vie Lognormale) et le cas de chocs (loi de durée de vie Weibull). Nous avons mis en place des outils de sélection de modèles permettant à l'ingénieur de réaliser ses études de fiabilité sans *a priori* sur les modèles d'accélération et d'exploiter les données issues d'essais accélérés en sur-stress.

Mots-clés : Fiabilité de semi-conducteurs, test accéléré, AFT, sélection de modèles, critère des moindres carrés pénalisés, modèles de régression définis par morceaux, loi de durée de vie Lognormale et Weibull, méthodes de calibration, détection de plages d'accélération en sur-stress.

Abstract : This thesis deals with the using of accelerating data and regression model selection for high technology field: semiconductor chips. The accelerating trail gives us regression frameworks. The aim of the accelerating test consists on fitting the logarithm of the lifetime through the use of some function f , called the acceleration function. However, accelerating data may have misleading and complex comportment. In order to adapt the model with such data, we have proposed to detect the changes on the comportment of the acceleration function. We have considered a collection of piecewise acceleration models candidate to the estimation. For each model candidate we have estimated the least-squares estimation. And we have selected the final estimator \tilde{f} using a penalized criterion. The penalized estimator is optimal approximation of the reality since the quadratic risk of penalized estimator is bounded by the minimal risk upon every least-squares estimators candidates. Moreover, this oracle inequality is non asymptotic. Furthermore, we have considered classical reliability cases: the Lognormal case associating with some fatigue failure, and the Weibull case associating with some choc failure. Lastly we have implemented model selection tools in order to realise survey study without *a priori* on the acceleration models and to use overstress trials.

Keywords : Semiconductor's reliability, accelerating test, AFT, model selection, penalized least-squares criterion, Lognormal and Weibull lifetimes, piecewise regression model, calibration procedure, detection of accelerating sets on overstress data.