

Ajustement de modèles probabilistes complexes et estimation des intervalles de confiance de manière discrète

Anaïs Tanguy, Sébastien Bosse
Centre National d'Etudes Spatiales CNES)
18, avenue Edouard Belin - 31401 Toulouse
Anais.tanguy@cnes.fr -
Sebastien.bosse@cnes.fr

André CABARBAYE
CNES / CAB INNOVATION
3 rue de la Coquille – 31500 Toulouse
andre.cabarbaye@cnes.fr /
andre.cabarbaye@cabinnovation.fr

Résumé

Le but de cet article est de montrer que des modèles probabilistes complexes peuvent être correctement ajustés à partir de données de retour d'expérience, sous réserve d'utiliser une méthode d'optimisation performante permettant de s'affranchir des multiples optima de la fonction de vraisemblance.

Il propose, par ailleurs, une méthode générique d'estimation des intervalles de confiance des paramètres et des quantiles, par inversion de la matrice de Fischer, calculées de manière discrète afin d'éviter la dérivation d'expressions analytiques propres à chacun des modèles.

Summary

The purpose of this communication is to demonstrate that relatively complex probabilistic models can be properly adjusted from operational data feedback, if a good optimization method is used.

This article is illustrated by many examples treated by a tool based on Genetic Algorithms and nonlinear Simplex.

It focuses on reliability multi-phase models, as Bertholon, multiple and simultaneous degradations models, acceleration models associated with different probability laws and models of preventive and corrective maintenance with possible phenomena of rejuvenation (GRP, Jack ...).

It also demonstrates the feasibility of the overall fit of predictive reliability models, as the FIDES guide, regarding the failure rates and acceleration factors.

Introduction

Divers modèles paramétriques sont proposés par les théoriciens de la Sécurité de Fonctionnement pour traduire le fonctionnement des systèmes. Ces modèles de probabilité sont très satisfaisants à l'esprit mais ne peuvent représenter la réalité que si leurs paramètres ont été correctement ajustés à partir de données de retour d'expérience.

La méthode d'ajustement la plus couramment utilisée est la méthode du maximum de vraisemblance qui consiste à rechercher la configuration de paramètres qui maximise la densité de probabilité obtenue par le modèle pour les valeurs des données expérimentales (ou la fiabilité pour les données censurées). Mais cette méthode ne peut se suffire d'une simple technique d'optimisation locale, telle que le pseudo gradient ou le simplexe non linéaire (algorithme de Nelder Mead), quand les optima sont multiples.

C'est la raison pour laquelle la plupart des outils existants se limite aux lois de probabilité usuelles et donne parfois des résultats erronés, qu'il s'agisse de l'ajustement d'une simple loi de Weibull à trois paramètres ou d'un modèle de fiabilité plus sophistiqué prenant en compte les conditions d'utilisation et d'environnement supportées par un matériel ou le type de maintenance dont il bénéficie. D'autres modèles particulièrement complexes, tels ceux proposés dans les guides de fiabilité prévisionnelle, ne sont pas véritablement confrontés au terrain et ne résultent que d'ajustements partiels de données constructeur et de jugements d'experts.

Bien que l'optimalité des solutions qu'elles procurent ne puisse être garantie ou démontrée, les méthodes stochastiques de type heuristique ou méta heuristique ont déjà largement prouvé leur efficacité dans la recherche d'optima globaux. Aussi, apparaît-il judicieux de les utiliser pour réaliser des ajustements, d'autant qu'elles ne nécessitent aucune connaissance a priori sur les fonctions à traiter, hormis le résultat obtenu pour chaque configuration de variables.

Aussi avons-nous tenté d'ajuster des modèles relativement complexes au moyen d'un outil d'optimisation globale (GENCAB de la société CAB INNOVATION). Nous nous sommes d'abord intéressés à des modèles de fiabilité multi-phases considérant globalement la courbe de taux de défaillance en baignoire (jeunesse, vie utile, usure), tel que celui de Bertholon [Ziani], ainsi qu'à la modélisation de phénomènes de dégradation multiples agissant simultanément sur un produit.

Nous avons ensuite traité plusieurs modèles d'accélération (Cox, Arhénus, Basquin...) associés à diverses lois de probabilité (exponentielle, Weibull, lognormale...) et démontré, par la même, la possibilité d'exploiter globalement des données de retour d'expérience issues de matériels soumis à des conditions diverses d'utilisation et d'environnement. Ainsi est-il envisageable d'enrichir un recueil de fiabilité prévisionnelle, tel que le guide FIDES [Groupe FIDES], tant en ce qui concerne les taux de défaillance de base que la valeur des paramètres entrant dans l'expression des facteurs d'accélération.

Nous nous sommes enfin intéressés à divers modèles de maintenance préventive et corrective qui considèrent d'éventuels phénomènes de rajeunissement (GRP, Jack [Jack]...).

Quand cela était possible, les résultats obtenus ont été comparés à ceux issus d'autres outils, en terme de vraisemblance calculée à partir d'un même jeu de test. A défaut, un jeu de données a été simulé à partir d'une configuration de paramètres du modèle puis un ajustement a été réalisé pour tenter de retrouver cette configuration de paramètres.

1 L'outil GENCAB

L'outil GENCAB [Cabarbaye] est basé sur une méthode hybride associant Algorithmes Génétiques, Evolution Différentielle et Simplexe non linéaire (algorithme de Nelder Mead). Son principe est illustré en figure 1. Constitués de divers paramètres (gènes) de type réel, entier ou binaire, les chromosomes sont aléatoirement soumis à des mutations, croisements et évolutions différentielles (somme d'un gène de chromosome avec la différence entre les mêmes gènes de deux autres chromosomes) puis, après sélection, les meilleurs éléments de la population bénéficient d'une amélioration au niveau local par un certain

nombre de pas de Simplexe. Cette hybridation de différentes techniques de recherche globale et locale, qui peuvent être éventuellement paramétrées par l'utilisateur, a pour finalité de rendre l'outil plus efficace et robuste face à la diversité des problématiques rencontrées. Ainsi l'Evolution Différentielle sera plus rapide que les Algorithmes Génétiques pour traiter des fonctions à peu près convexes mais présentera l'inconvénient, pour d'autres, de jouer simultanément sur l'ensemble des gènes.

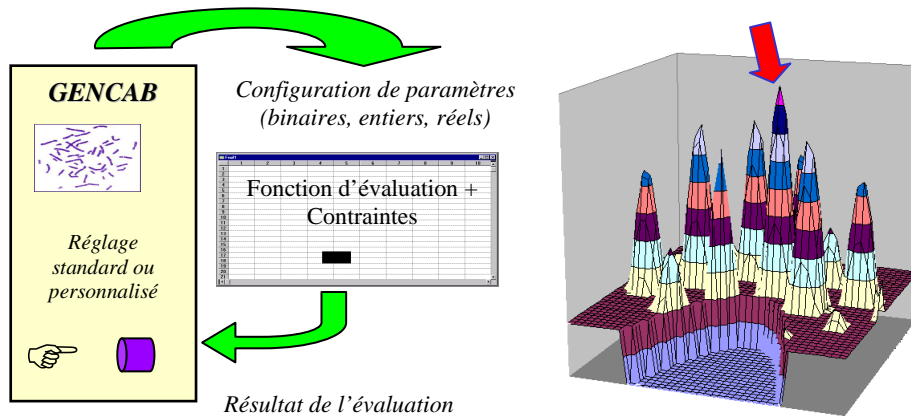


Figure 1. Principe de l'outil d'optimisation GEN CAB

2 Modèles de fiabilité

2.1 Modèles de fiabilité multi-phases

Le modèle de BERTHOLON associe une exponentielle à une Weibull pour modéliser globalement la deuxième et la troisième partie de la courbe en baignoire (défaillances occasionnelles et usure). La fiabilité peut s'exprimer par la formule 1, la densité de probabilité par la formule 2 et l'occurrence d'une défaillance peut être simulée sous Excel par la formule 3.

$$R(t) = \exp(-\lambda t) * \exp(-[\max(0, t-T)/\eta]^\beta) \quad (1)$$

$$f(t) = \lambda(t) * R(t) \text{ avec } \lambda(t) = \lambda + SI(t > T; \beta * (t-T)^{\beta-1} / \eta^\beta; 0) \quad (2)$$

$$\text{MIN}(-\text{LN}[ALEA()]/\lambda; T + \eta * (-\text{LN}[ALEA()])^{1/\beta}) \quad (3)$$

La figure 2 présente un ajustement du modèle de Bertholon réalisé à partir de 100 valeurs préalablement simulées. Les courbes correspondent aux fonctions de répartition du modèle théorique, du modèle expérimental (données simulées) et du modèle ajusté. On retrouve approximativement les valeurs des 4 paramètres du modèle utilisé pour la simulation.

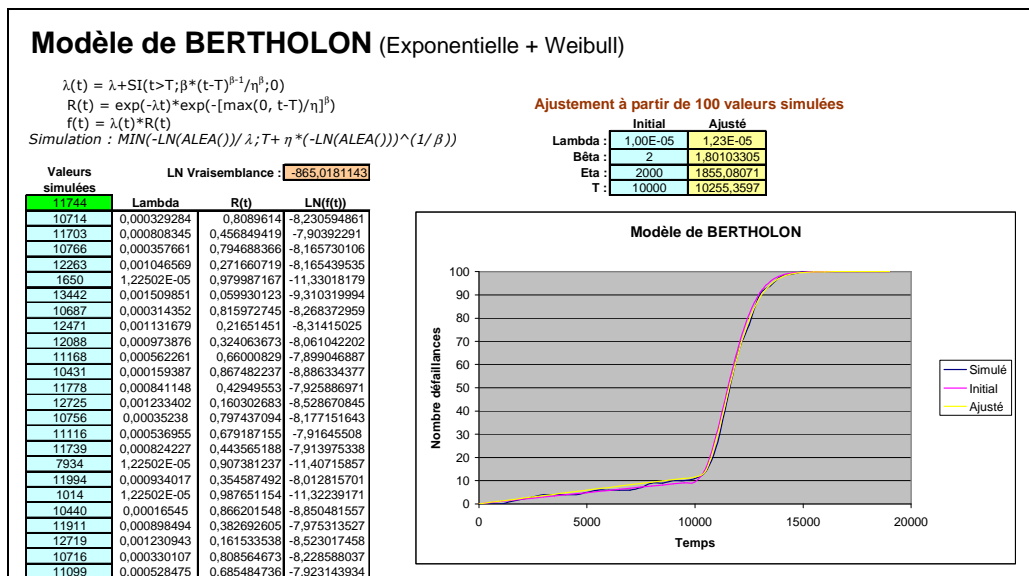


Figure 2. Ajustement du modèle de Bertholon à partir de données simulées

Ce modèle peut être généralisé en un modèle à 7 paramètres caractérisant les trois phases de la courbe en baignoire, soit une première Weibull à $\beta < 1$ pour la phase de jeunesse, une exponentielle pour la phase à taux constant et une seconde Weibull à $\beta > 1$ pour la phase de vieillissement à taux croissant. La fiabilité peut alors s'exprimer par la formule 4, la densité de probabilité par la formule 5 et l'occurrence d'une défaillance par la formule 6.

$$R(t) = \exp(-[\min(t; T_1)/\eta_1]^{\beta_1}) * \exp(-\lambda t) * \exp(-[\max(0; t-\gamma_2)/\eta_2]^{\beta_2}) \quad (4)$$

$$f(t) = \lambda(t) * R(t) \text{ avec } \lambda(t) = \text{si}(t < T_1; \beta_1 * t^{\beta_1-1} / \eta_1^{\beta_1}; 0) + \lambda + \text{si}(t > \gamma_2; \beta_2 * (t-\gamma_2)^{\beta_2-1} / \eta_2^{\beta_2}; 0) \quad (5)$$

$$T_i = \eta_1 * (-\text{LN}[ALEA()])^{1/\beta_1} \quad T = \text{si}(T_i < T_1; T_i; \text{MIN}(-\text{LN}[ALEA()]/\lambda; \gamma_2 + \eta_2 * (-\text{LN}[ALEA()])^{1/\beta_2}) \quad (6)$$

La figure 3 présente l'ajustement d'un tel modèle réalisé à partir de 100 valeurs préalablement simulées. A nouveau, on retrouve approximativement les valeurs des paramètres du modèle utilisé pour la simulation.

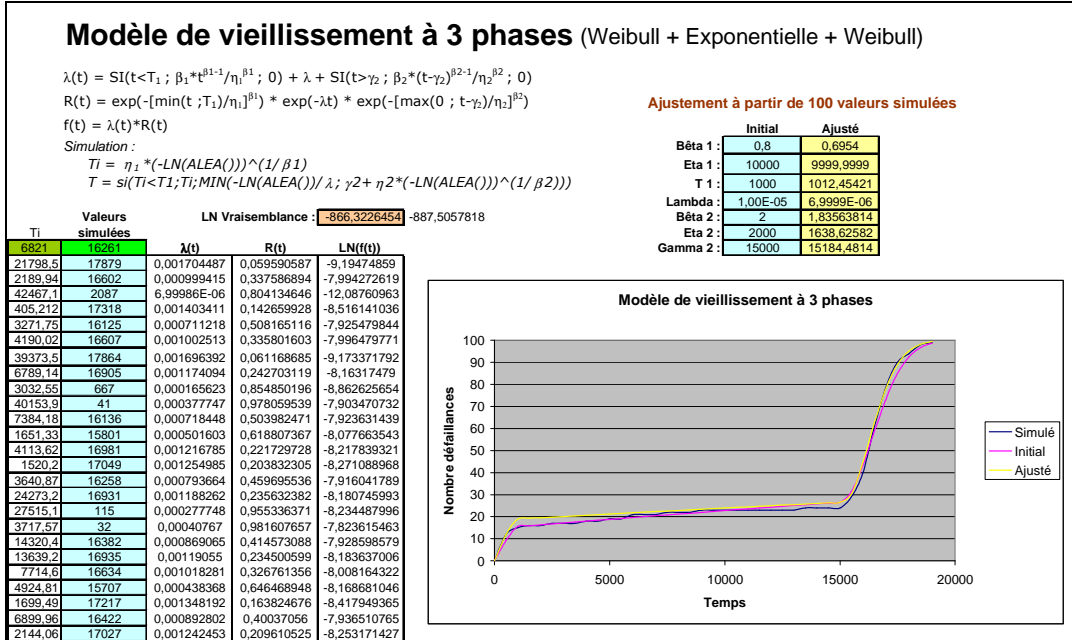


Figure 3. Ajustement d'un modèle à 3 phases à partir de données simulées

2.2 Modèles de dégradation multiples

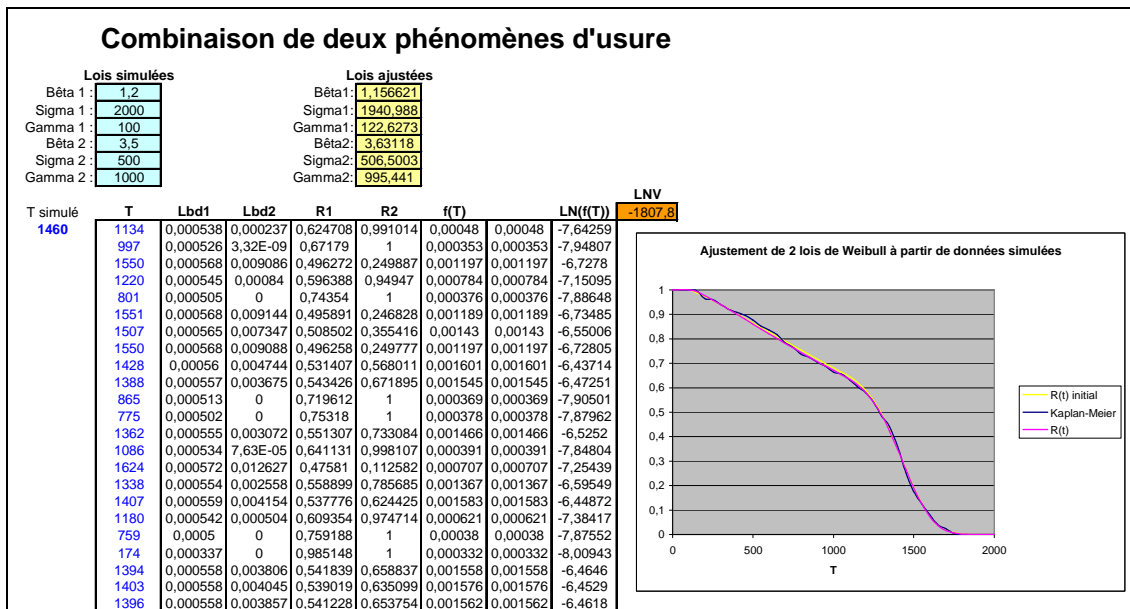


Figure 4. Ajustement de deux lois de Weibull relatives à des phénomènes d'usure concomitants

Dans le cas d'un équipement soumis simultanément à deux phénomènes d'usure indépendants modélisés chacun par une loi de Weibull, la fiabilité peut s'exprimer par la formule 7, la densité de probabilité par la formule 8 et l'occurrence d'une défaillance par la formule 9. La figure 4 présente l'ajustement d'un tel modèle réalisé à partir de 250 valeurs simulées sans aucun classement préalable de celles-ci. On retrouve approximativement les paramètres des deux lois de Weibull ayant servi à simuler le jeu de données.

$$R(t) = R_1(t) * R_2(t) = \exp(-[\max(0 ; t-\gamma_1)/\eta_1]^{\beta_1}) * \exp(-[\max(0 ; t-\gamma_2)/\eta_2]^{\beta_2}) \quad (7)$$

$$f(t) = \lambda(t) * R(t) \quad \text{avec } \lambda(t) = \lambda_1(t) + \lambda_2(t) = \beta_1 \max(0 ; t-\gamma_1)^{\beta_1-1} / \sigma_1^{\beta_1} + \beta_2 \max(0 ; t-\gamma_2)^{\beta_2-1} / \sigma_2^{\beta_2} \quad (8)$$

$$T_i = \text{MIN}(\gamma_1 + \sigma_1 * (-\text{LN}(\text{ALEA}()))^{1/\beta_1}; \gamma_2 + \sigma_2 * (-\text{LN}(\text{ALEA}()))^{1/\beta_2}) \quad (9)$$

3 Modèles d'accélération

Les modèles de vieillissement accéléré sont utilisés pour réduire la durée des essais de fiabilité en renforçant les conditions d'utilisation ou d'environnement. Mais ils peuvent être également utilisés pour ajuster des modèles de fiabilité à partir de données opérationnelles obtenus sous diverses conditions, comme le montrent les exemples suivants.

3.1 Modèle d'Arrhenius couplé à une loi de Weibull

Le facteur d'accélération d'Arrhenius permet de faire correspondre des durées de fonctionnement équivalentes à des températures différentes

$$FA = e^{\frac{Ea}{K} \left(\frac{1}{T_1} - \frac{1}{T_2} \right)} \quad \text{Avec } T \text{ la température en Kelvin, } Ea \text{ l'énergie d'activation et } K \text{ la constante de Boltzmann.} \quad (10)$$

Quand elle est inconnue, l'énergie d'activation peut être considérée comme un paramètre supplémentaire à ajuster comme dans l'exemple de la figure 5 traitant d'une loi de Weibull.

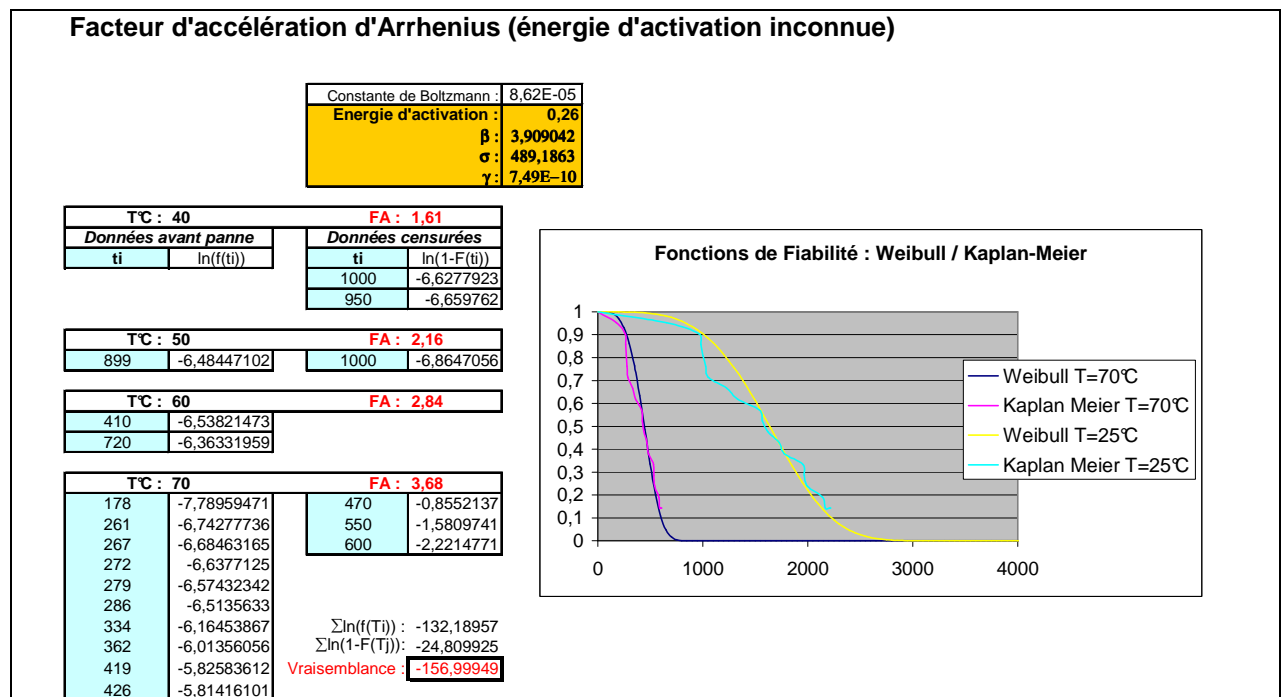


Figure 5. Ajustement d'un modèle d'Arrhenius couplé à une loi de Weibull

3.2 Modèle FIDES

Le guide FIDES propose des modèles de fiabilité relativement complexes basés sur le modèle de Cox [Cox] en considérant différents contributeurs de stress : la température, le cyclage thermique, l'humidité, le niveau de vibration et la tension électrique. Ce guide décompose le taux de défaillance physique d'un composant en taux de défaillance élémentaires sur lesquels portent certains facteurs d'accélération :

$$\lambda_{\text{Physique}} = \lambda_{\text{OTH}} \prod_{\text{Thermique}} + \lambda_{\text{Boîtier}} \prod_{\text{Boîtier}} + \lambda_{\text{Joints brasés}} \prod_{\text{Joints brasés}} + \lambda_{\text{Méca}} \prod_{\text{Méca}} + \lambda_{\text{ORH}} \prod_{\text{ORH}} \quad (11)$$

- Une loi d'Eyring est utilisée pour la puce ($\prod_{\text{Thermique}}$),
- Une loi de Norris Landzberg est utilisée pour le boîtier ($\prod_{\text{Boîtier}}$) ; celle-ci étant modifiée pour passer d'un nombre de cycles à une durée,
- Une seconde loi de Norris Landzberg est utilisée pour les joints brasés ($\prod_{\text{Joints brasés}}$),
- Une loi de Basquin est utilisée pour la mécanique ($\prod_{\text{Méca}}$),
- Une loi de Peck est utilisée pour l'humidité (\prod_{ORH}).

Cette décomposition a pour effet de multiplier le nombre de paramètres du modèle et complique singulièrement, par la-même, son ajustement à partir de données de retour d'expérience. Un tel ajustement a cependant été réalisé à partir de données de défaillance simulées dans des environnements et conditions d'utilisation hétérogènes, afin de tenter de retrouver les différents paramètres du modèle et de démontrer, par là-même, sa faisabilité. Comme l'illustre la figure 6, un jeu de 400 durées de défaillance a été simulé à partir de conditions environnementales et d'utilisation tirées uniformément entre des valeurs min et max. Connaissant les taux de défaillance λ_{0i} et les paramètres entrant dans les facteurs d'accélération, le taux de défaillance λ est d'abord calculé puis une durée avant défaillance est obtenue par tirage aléatoire entre 0 et 1 et application de l'inverse de la fonction $\exp(-\lambda t)$.

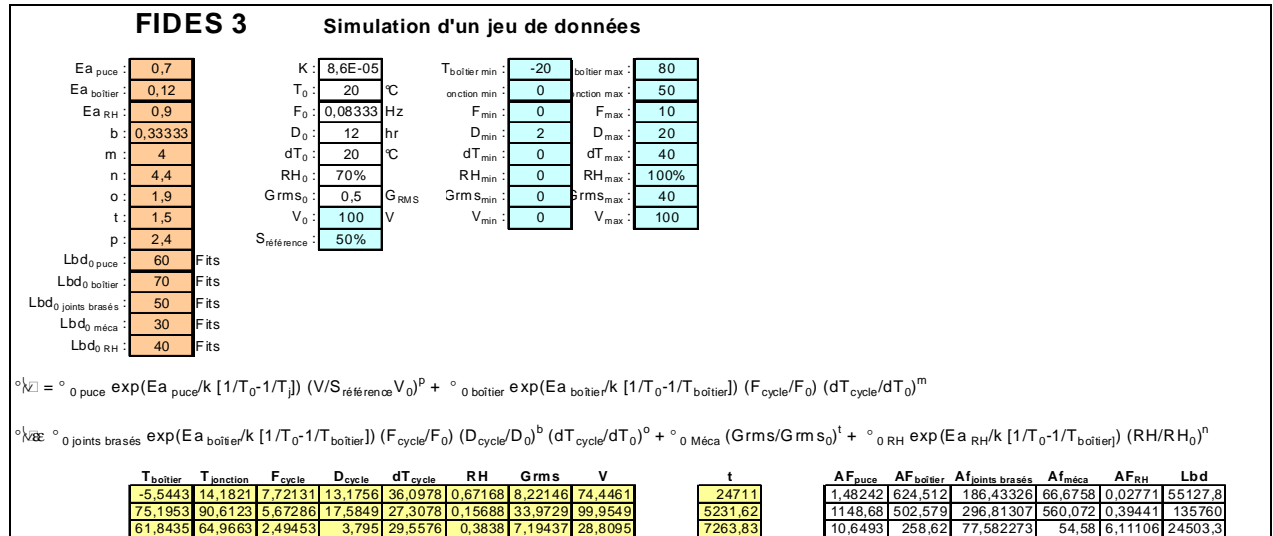


Figure 6. Simulation du modèle FIDES

Comme le montre la figure 7, l'ajustement des 14 paramètres du modèle est possible et la vraisemblance obtenue est supérieure à celle calculée avec les paramètres initiaux.

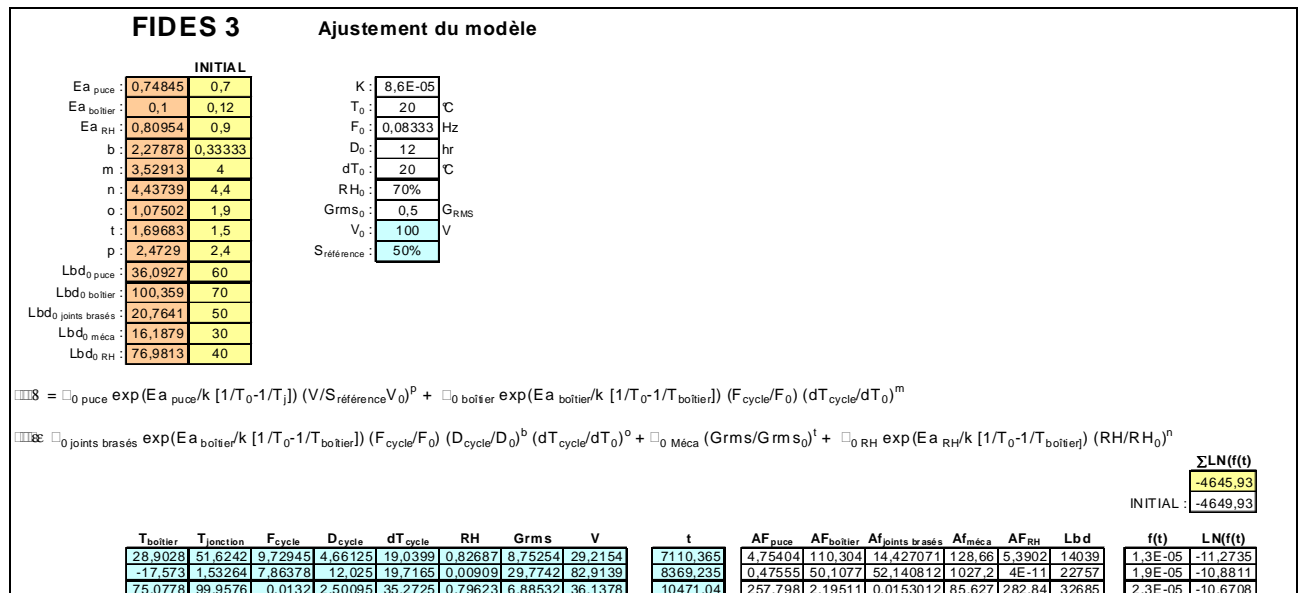


Figure 7. Ajustement du modèle FIDES

Le grand nombre de paramètres du modèle et sa sensibilité au bruit de simulation rendent toutefois délicat son ajustement global qui nécessite beaucoup de données, ce qui n'est évidemment plus le cas si les modes de défaillance des composants défectueux peuvent être identifiés précisément.

4 Modèles de maintenance

4.1 Processus de remplacement généralisé GRP (Generalized Renewal Process)

Après une maintenance corrective, l'état d'un équipement peut être aussi bon que neuf (as good as new), dans l'état correspondant à son âge (as bad as old) ou dans un état intermédiaire (better than old but worse than new), en excluant les états mieux que neuf ou pire que son âge.

Aussi 3 modèles ont-ils été proposés pour modéliser respectivement ces 3 types de maintenance :

- Le processus de remplacement RP (Renewal Process)
- Le processus non homogène de Poisson NHPP (Non-homogeneous Poisson process)
- Le processus de remplacement généralisé GRP (Generalized Renewal Process) de type 1 ou 2.

Dans le modèle GRP type 1, la maintenance corrective a un effet de rajeunissement de l'équipement proportionnel à la durée écoulée depuis la maintenance précédente. L'âge virtuel de l'équipement juste après la rⁱème action de maintenance peut s'exprimer par la formule 12.

$$A_r = A_{r-1} + q^*(t_r - t_{r-1}) = A_{r-2} + q^*(t_{r-1} - t_{r-2}) + q^*(t_r - t_{r-1}) = q^r t_r \text{ avec } q \text{ le facteur de rajeunissement (complet : } q = 0 \text{ nul : } q = 1) \quad (12)$$

La probabilité que l'équipement soit en panne à t, sachant qu'il a été réparé à l'instant t_r est donnée par la formule 13.

Si le vieillissement de l'équipement est modélisé par une loi de Weibull, la fonction de répartition correspondante est donnée par la formule 14 et celle de la densité de probabilité par la formule 15.

$$[F(t)-F(t_r)]/R(t_r) = (1-R(t)-1+R(t_r)) / R(t_r) = 1 - R(t)/R(t_r) \quad (13)$$

$$F(t) = 1 - \exp[-(qt_r/\sigma)^\beta - ((qt_r+(t-t_r))/\sigma)^\beta] \quad (14)$$

$$f(t) = \beta(qt_r+(t-t_r))^{\beta-1}/\sigma^\beta \exp[-(qt_r/\sigma)^\beta - ((qt_r+(t-t_r))/\sigma)^\beta] \quad (15)$$

Le modèle GRP type 2 diffère du modèle GRP type 1 par le fait que la maintenance corrective conduit à un rajeunissement de l'équipement proportionnel à son âge virtuel. Les expressions précédentes sont alors données par les formules 16, 17 et 18.

$$A_r = q^*(A_{r-1} + t_r - t_{r-1}) = q^*(q^*(A_{r-2} + t_{r-1} - t_{r-2}) + t_r - t_{r-1}) = q^r * t_1 + q^{r-1} * (t_2 - t_1) + \dots + q * (t_r - t_{r-1}) = \sum_{i=1}^r q^{r-i+1} * (t_i - t_{i-1}) \quad (16)$$

$$F(t) = 1 - \exp[-(Ar/\sigma)^\beta - ((Ar+t-t_r)/\sigma)^\beta] \quad (17)$$

$$f(t) = \beta(Ar+t-t_r)^{\beta-1}/\sigma^\beta \exp[-(Ar/\sigma)^\beta - ((Ar+t-t_r)/\sigma)^\beta] \quad (18)$$

Dans le cadre d'une étude comparative réalisée sur ce type de modèle à partir de différents jeux de données, les ajustements réalisés par cet outil d'optimisation globale ont toujours présenté une vraisemblance égale ou supérieure aux autres. Nous éviterons, toutefois, de présenter ici les résultats aberrants fournis par certains outils largement diffusés.

La figure 8 présente l'ajustement d'un modèle GRP type 2 réalisé à partir de 200 valeurs simulées. On retrouve approximativement les 3 paramètres du modèle.

Processus de maintenance GRP 2

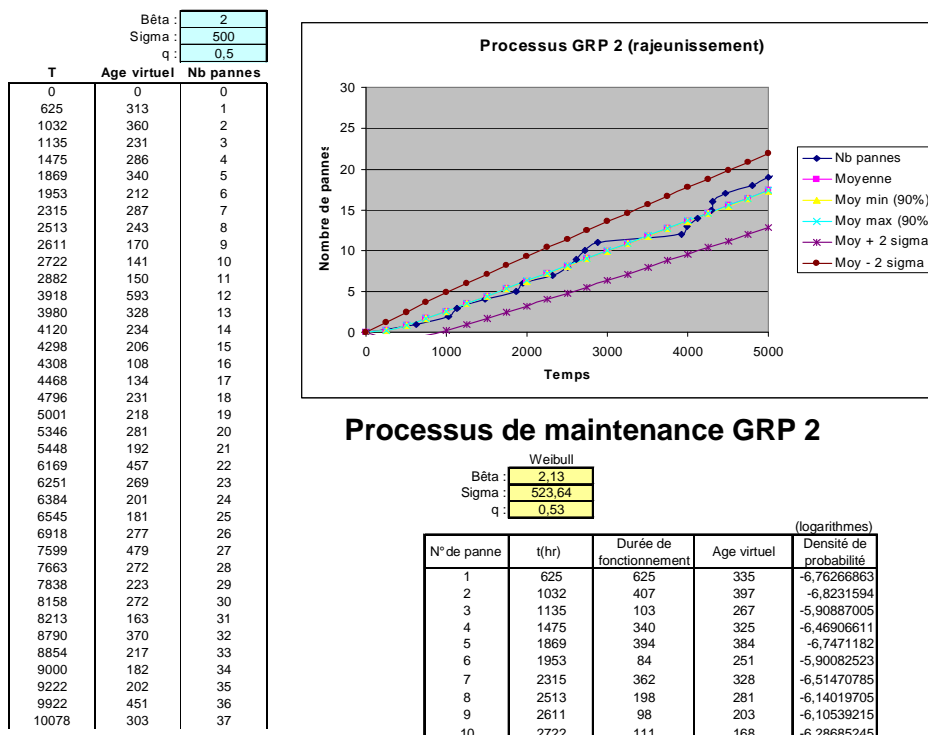


Figure 8. Ajustement d'un modèle GRP type 2 à partir de données simulées

4.2 Prise en compte de la maintenance préventive (modèle de Jack)

Jack a proposé plusieurs modèles de vieillissement dans lesquels la maintenance a un effet de rajeunissement de l'équipement. Cet effet est plus important dans le cas d'une action de maintenance préventive (changement de plusieurs pièces d'usure) que corrective (changement de l'unique pièce en panne). Dans le modèle Jack type 2, l'âge virtuel d'un équipement à la fin d'une

action de maintenance est égal à celui qu'il avait juste avant celle-ci, multiplié par une proportion p_p dans le cas d'une maintenance préventive ou p_c dans le cas d'une maintenance corrective.
De la même manière que pour le modèle GRP, le modèle Jack type 2 a été simulé en considérant une maintenance préventive périodique puis ajusté à partir d'un échantillon de 200 valeurs simulées. On retrouve à nouveau les 4 paramètres utilisés pour la simulation comme le montre la figure 9.

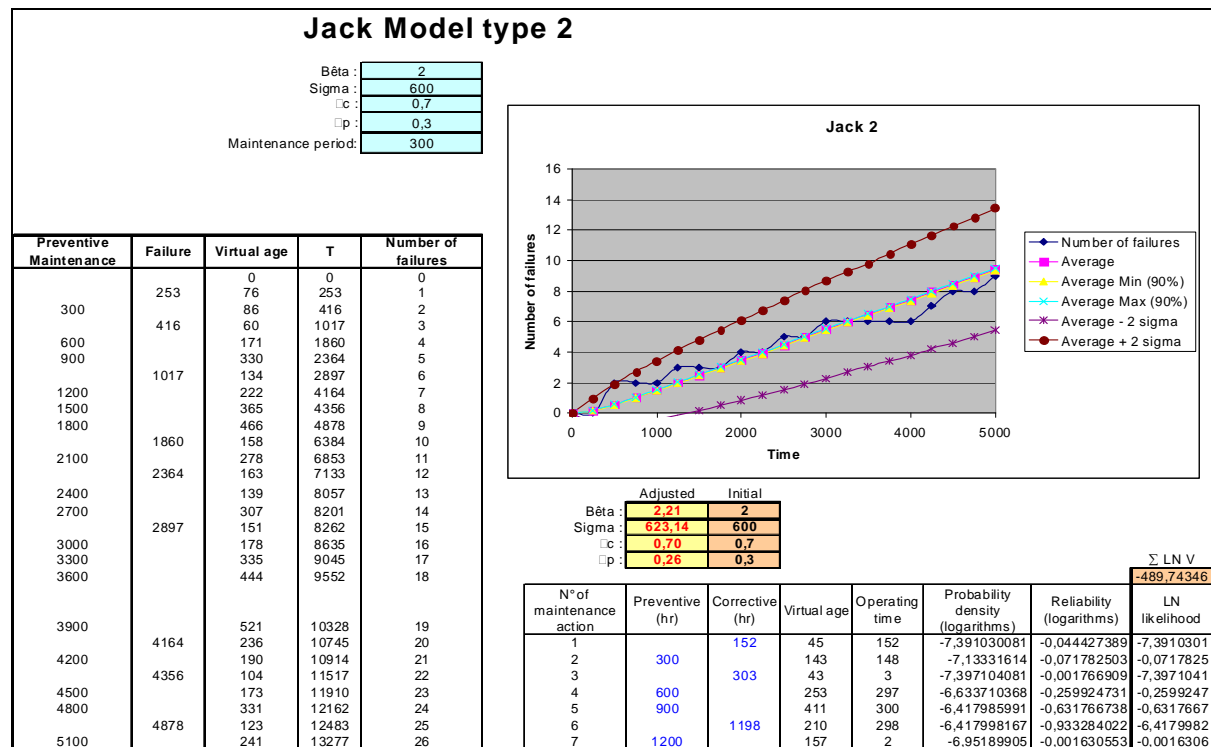


Figure 9. Ajustement d'un modèle Jack type 2 à partir de données simulées

CONCLUSION

Un outil d'optimisation globale s'avère indispensable à l'ajustement de modèles paramétriques d'une certaine complexité. Il facilite l'exploitation de données de retour d'expérience et permet notamment d'agréger des données issues d'environnement et conditions d'utilisation divers, ou relatives à différents modes de défaillance, sans faire d'hypothèse a priori sur des facteurs de recalage ou de tri préalable des données.
Il pourrait ainsi permettre d'enrichir et d'actualiser des recueils de fiabilité prévisionnelle, tel que le guide FIDES, tant en ce qui concerne les taux de défaillance de base que la valeur des paramètres entrant dans l'expression des facteurs d'accélération.
Il rend également possible l'utilisation de modèles relativement sophistiqués, tel que celui que nous avons proposé pour caractériser les trois phases de la courbe en baignoire d'un taux de défaillance.
Dans le domaine de la maintenance, il permet de mieux suivre l'exploitation des matériels et d'effectuer des optimisations sur des paramètres divers tels que la durée d'amortissement ou l'occurrence optimale des actions de maintenance préventive.
Basé sur l'outil d'optimisation utilisé ici (GENCAB), un outil spécifique à l'ajustement de modèles complexes (FITCAB) est en cours de développement.

Références

- [1] Ziani, R. & Antoni, M. (2008). Modélisation du vieillissement des appareils de signalisation par le modèle de Bertholon et optimisation de la maintenance, lambda mu 16, 7-9 octobre 2008, Avignon, France.
- [2] Cabarbaye, A. (2003). Outil générique d'optimisation dans le domaine discret et/ou continu éventuellement stochastique - ROADEF'03 – Avignon.
- [3] Jack, N. (1998). Age-reduction models for imperfect maintenance. IMA Journal of Management Mathematics 9, pp. 347-354.
- [4] Cox, D. R. (1972). "Regression Models and Life Tables". Journal of the Royal Statistical Society Series B 34 (2): 187–220.
- [5] Groupe FIDES. "Méthodologie de fiabilité pour les systèmes électroniques". Guide FIDES 2009.