

Apport d'un outil d'optimisation globale à l'ajustement des modèles statistiques

André Cabarbaye, Julien Faure

► **To cite this version:**

André Cabarbaye, Julien Faure. Apport d'un outil d'optimisation globale à l'ajustement des modèles statistiques. 41èmes Journées de Statistique, SFdS, Bordeaux, 2009, Bordeaux, France, France. 2009. <inria-00386738>

HAL Id: inria-00386738

<https://hal.inria.fr/inria-00386738>

Submitted on 22 May 2009

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

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

APPORT D'UN OUTIL D'OPTIMISATION GLOBALE A L'AJUSTEMENT DES MODELES STATISTIQUES

CABARBAYE André ^{1&2}, FAURE Julien ¹

¹ *Centre National d'Etudes Spatiales (CNES), 18 av. Edouard Belin, 31401, Toulouse, France
05 61 28 27 41 / 05 61 27 47 19, andre.cabarbaye@cnes.fr / julien.faure@cnes.fr*

² *CAB INNOVATION, 3 rue de la Coquille, 31500, Toulouse, France
05 61 54 68 08, andre.cabarbaye@cabinnovation.fr*

Résumé : Les modèles paramétriques 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 utilisée est la méthode du maximum de vraisemblance qui ne peut se suffire d'une simple technique d'optimisation locale quand les optima sont multiples. C'est la raison pour laquelle les outils existants donnent parfois des résultats erronés. Aussi avons-nous tenté d'ajuster différents modèles utilisés en fiabilité au moyen d'un outil d'optimisation globale associant Algorithmes Génétiques, Evolution Différentielle et Simplexe non linéaire. Nous nous sommes d'abord intéressés aux processus de renouvellement généralisé (GRP), pour lesquels les actions de maintenance ont un certain effet de rajeunissement. Nous avons ensuite traité le modèle de BERTHOLON, utilisé dans le domaine ferroviaire, qui associe une exponentielle à une Weibull pour modéliser globalement la deuxième et la troisième partie de la courbe en baignoire (défaillances précoces, défaillances occasionnelles puis usure). Ce modèle est susceptible d'être largement utilisé à l'avenir pour caractériser la fiabilité des composants électroniques dont l'intégration croissante devrait conduire à des limitations en durée de vie de plus en plus sévères. Mais son ajustement présentait jusqu'alors des difficultés. Pour considérer globalement les trois phases de la courbe en baignoire, nous avons enfin proposé un modèle plus complet à 7 paramètres que nous avons pu également ajuster sans difficulté en retrouvant approximativement les paramètres utilisés pour simuler préalablement un jeu d'une centaine de données expérimentales.

Summary : The parametric models can be used only if their parameters have been correctly adjusted from data feedback. The maximum likelihood method is the adjustment method used most of the time. But it is not possible to use a simple local optimization technique when optima are numerous. That is why the existing tools may give erroneous results. So we tried to adjust different models used in the reliability domain through a global optimization tool combining Genetic Algorithms, Differential Evolution and Nonlinear Simplex. We are first interested in the generalized renewal process (GRP), for which maintenance actions have a rejuvenation effect. Then, we treated the Bertholon model, used in the railway domain, which combines an exponential to a Weibull to model overall the second and third parts of the bathtub curve (early failures, occasional failures, wear failures). This model is likely to be widely used in the future to characterize the reliability of electronic components whose integration should lead to more and more severe limitations in life duration. But its adjustment was difficult before. To model globally the three phases of the bathtub curve, we have finally proposed a more comprehensive model with 7 parameters that we have been able to adjust without difficulty finding approximately the parameters used before to simulate a sample of one hundred experimental data.

Mots-clés : Qualité – fiabilité, Statistique mathématique

1- Introduction

De nombreux modèles paramétriques sont proposés dans les divers domaines concernés par les statistiques tels que la fiabilité ou la finance par exemple. Ces modèles théoriques sont souvent très satisfaisants à l'esprit, tant la logique qui les sous-tend apparaît rationnelle. Mais ils ne peuvent représenter la réalité que si leurs paramètres ont été correctement ajustés à partir de données de retour d'expérience.

La méthode la plus couramment utilisée pour ajuster un modèle probabiliste est la méthode du maximum de vraisemblance qui consiste à rechercher la configuration de paramètres qui maximise la densité de probabilité donnée par le modèle pour les valeurs expérimentales (produit des densités ou somme de leurs logarithmes). Mais cette méthode peut difficilement se suffire d'une simple technique d'optimisation locale, tel 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 majorité des outils existants donnent parfois des résultats erronés notamment quand le nombre des paramètres du modèle considéré est supérieur à deux. Cela peut être le cas pour une simple Weibull dont le paramètre γ est non nul ou un modèle de fiabilité plus sophistiqué censé représenter le comportement d'un produit en fonction de la sévérité de son environnement (modèle d'accélération des défaillances) ou du type de maintenance dont il bénéficie.

Aussi avons-nous tenté d'ajuster différents modèles utilisés en fiabilité au moyen d'un outil d'optimisation globale (GENCAB de la société CAB INNOVATION) basé sur une méthode hybride associant Algorithmes Génétiques, Evolution Différentielle et Simplexe non linéaire.

Nous nous sommes d'abord intéressés aux processus de renouvellement généralisé (GRP) [1] dans lesquels les actions de maintenance ont un certain effet de rajeunissement du produit (better than old but worse than new).

Nous avons ensuite traité le modèle de BERTHOLON [2], utilisé dans le domaine ferroviaire, qui associe une exponentielle et une Weibull pour modéliser globalement la deuxième et la troisième partie de la courbe en baignoire (défaillances précoces, occasionnelles puis usure). Ce modèle, dont l'ajustement posait apparemment problème jusqu'alors [2], nous semble susceptible d'être largement utilisé à l'avenir pour caractériser la fiabilité des composants électroniques dont l'intégration croissante devrait conduire à des limitations en durée de vie de plus en plus sévère.

Pour considérer globalement les trois phases de la courbe en baignoire, nous avons enfin proposé un modèle plus complet à 7 paramètres que nous avons pu également ajuster sans difficulté en retrouvant approximativement les paramètres utilisés pour simuler préalablement un jeu d'une centaine de données expérimentales.

2 – L'outil GENCAB

L'outil GENCAB [3] est basé sur une méthode hybride associant Algorithmes Génétiques [4], Evolution Différentielle [5] 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 robuste à la diversité des problématiques rencontrées. Ainsi l'Evolution Différentielle sera généralement plus efficace que les Algorithmes Génétiques pour traiter des fonctions convexes mais présentera l'inconvénient, pour d'autres, de jouer simultanément sur l'ensemble des gènes.

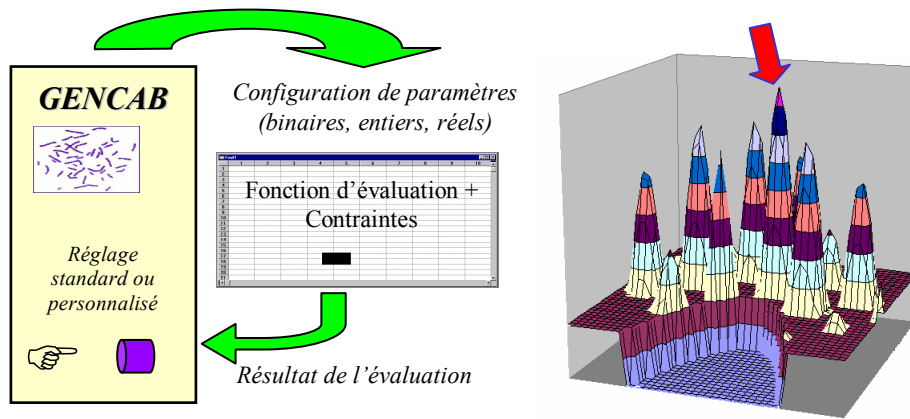


Figure 1. Principe de l'outil d'optimisation GENCAB

3 – 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 A_r de l'équipement à l'instant t_r juste après la r ième action de maintenance est égal à :

$$A_r = A_{r-1} + q*(t_r - t_{r-1}) = A_{r-2} + q*(t_{r-1} - t_{r-2}) + q*(t_r - t_{r-1}) = q*t_r$$

avec q le facteur de rajeunissement.

$q = 0$: rajeunissement complet (équivalent à un processus RP)

$q = 1$: absence de rajeunissement (équivalent à un processus NHPP)

La probabilité que l'équipement soit en panne à t , sachant qu'il a été réparé à t_r est égale à :

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

Si le vieillissement de l'équipement est modélisé par une loi de Weibull, la fonction de répartition correspondante est :

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

Et celle de la densité de probabilité est : $f(t) = \beta(qtr + (t - t_r))^{\beta-1} / \sigma^\beta \exp[-(qtr/\sigma)^\beta - ((qtr + (t - t_r))/\sigma)^\beta]$

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. L'âge virtuel A_r de l'équipement à l'instant t_r juste après la r ième action de maintenance est égal à :

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

$$A_r = \sum_{i=1}^r q^{r-i+1} * (t_i - t_{i-1})$$

La fonction de répartition correspondante est : $F(t) = 1 - \exp[-(A_r/\sigma)^\beta - ((A_r + t - t_r)/\sigma)^\beta]$

Et celle de la densité de probabilité est : $f(t) = \beta(A_r + t - t_r)^{\beta-1} / \sigma^\beta \exp[-(A_r/\sigma)^\beta - ((A_r + t - t_r)/\sigma)^\beta]$

Dans le cadre d'une étude comparative réalisée sur ce type de modèle à partir de différents jeux de données, les ajustements réalisés par l'outil GENCAB ont toujours présenté une variance égale ou

supérieure à celle obtenue avec d'autres outils (figure 2). Nous éviterons, cependant, de présenter ici les résultats fournis par certains outils largement diffusés.

Processus de maintenance GRP 1

Weibull				
Bêta :	2,09			
Sigma :	329,85			
q :	0,23			

(logarithmes)				
N° de panne	t(hr)	Durée de fonctionnement	Densité de probabilité	
1	243	243	-5,92383488	
2	360	117	-6,001509	
3	463	103	-5,93401074	
4	552	89	-5,87513565	
5	852	300	-6,3576689	
6	1320	467	-8,25007015	
7	1464	145	-5,78491128	
8	1543	78	-5,38730487	
9	1930	387	-8,42863938	
10	1956	27	-4,92190646	
11	2101	145	-5,92704921	
12	2219	118	-5,68464883	
13	2257	38	-4,91806641	
14	2331	74	-5,25022309	
15	2337	6	-4,59656746	
16	2339	3	-4,5628046	
17	2370	30	-4,81131336	
18	2481	112	-5,66643925	
19	2523	42	-4,89297751	
20	2594	71	-5,21366003	
			-114,388741	Max

Processus de maintenance GRP 2

Weibull				
Bêta :	1,66			
Sigma :	426,39			
q :	1,00			

(logarithmes)				
N° de panne	t(hr)	Durée de fonctionnement	Age virtuel	Densité de probabilité
1	243	243	243	-6,31311211
2	360	117	360	-6,02236289
3	463	103	463	-5,88680433
4	552	89	552	-5,7667371
5	852	300	852	-6,71223437
6	1320	467	1320	-8,16451474
7	1464	145	1464	-5,96452887
8	1543	78	1543	-5,40178674
9	1930	387	1930	-8,35240919
10	1956	27	1956	-4,82564984
11	2101	145	2101	-6,07118423
12	2219	118	2219	-5,79407359
13	2257	38	2257	-4,89348959
14	2331	74	2331	-5,3000978
15	2337	6	2337	-4,50186858
16	2339	3	2339	-4,4587774
17	2370	30	2370	-4,78068846
18	2481	112	2481	-5,75347525
19	2523	42	2523	-4,89964295
20	2594	71	2594	-5,26355104
				-115,126989

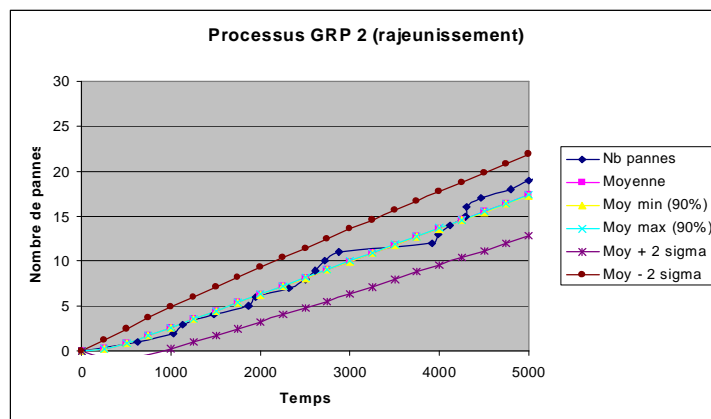
Figure 2. Ajustement de modèles GRP

De même, l'ajustement permet de retrouver approximativement les paramètres d'un modèle GRP utilisé pour simuler préalablement un jeu de données (figure 3).

Processus de maintenance GRP 2

Weibull		
Bêta :	2	
Sigma :	500	
q :	0,5	

T	Age virtuel	Nb pannes
0	0	0
625	313	1
1032	360	2
1135	231	3
1475	286	4
1869	340	5
1953	212	6
2315	287	7
2513	243	8
2611	170	9
2722	141	10
2882	150	11
3918	593	12
3980	328	13
4120	234	14
4298	206	15
4308	108	16
4468	134	17
4796	231	18
5001	218	19
5346	281	20
5448	192	21
6169	457	22
6251	269	23
6384	201	24
6545	181	25
6918	277	26
7599	479	27
7663	272	28
7838	223	29
8158	272	30
8213	163	31
8790	370	32
8854	217	33
9000	182	34
9222	202	35
9922	451	36
10078	303	37



Processus de maintenance GRP 2

Weibull				
Bêta :	2,13			
Sigma :	523,64			
q :	0,53			

(logarithmes)				
N° de panne	t(hr)	Durée de fonctionnement	Age virtuel	Densité de probabilité
1	625	625	335	-6,76266863
2	1032	407	397	-6,8231594
3	1135	103	267	-5,90887005
4	1475	340	325	-6,46906611
5	1869	394	384	-6,7471182
6	1953	84	251	-5,90082523
7	2315	362	328	-6,51470785
8	2513	198	281	-6,14019705
9	2611	98	203	-6,10539215
10	2722	111	168	-6,28685245

Figure 3. Simulation d'occurrences de pannes puis ajustement d'un modèle GRP (200 valeurs)

4 – Modèle de BERTHOLON

Le modèle de Bertholon correspond à la mise en série de deux blocs, le premier régi par une loi de mortalité exponentielle et le second par une loi de Weibull.

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

L'occurrence d'une défaillance peut être simulée en considérant la plus faible valeur simulée à partir de ces deux modèles (obtenue en inversant les fonctions de répartition ou de fiabilité pour une valeur tirée aléatoirement entre 0 et 1), soit :

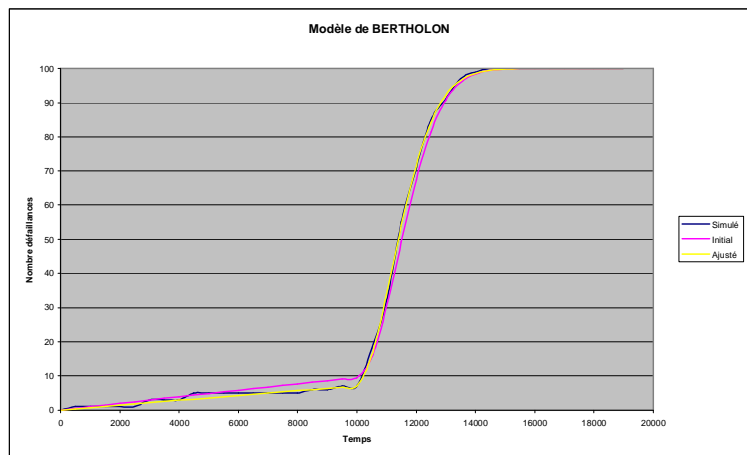
$$\text{MIN}(-\text{LN}[\text{ALEA}()]/\lambda ; T + \eta * (-\text{LN}[\text{ALEA}()])^{1/\beta}) \quad \text{sous Excel}$$

La densité de la loi de Bertholon peut s'exprimer de la manière suivante : $f(t) = \lambda(t) * R(t)$

avec $\lambda(t) = \lambda + \text{SI}(t > T ; \beta * (t-T)^{\beta-1} / \eta^\beta ; 0)$ et $R(t) = \exp(-\lambda t) * \exp(-[\max(0, t-T)/\eta]^\beta)$

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

Valeurs simulées	Lambda	R(t)	LN(f(t))
10503	0,000349677	0,838708505	-8,134392409
11717	0,000946266	0,378091446	-7,935605935
11290	0,000747039	0,542865757	-7,810286412
11810	0,000988673	0,345548304	-7,981769399
11865	0,001013198	0,327328131	-8,011435451
10720	0,000464344	0,767975416	-7,938882017
11123	0,000666228	0,611241014	-7,806142027
2900	7,30849E-06	0,979027915	-11,84766864
11570	0,000878485	0,43243491	-7,87563544
12111	0,00112384	0,251411265	-8,171669178
11472	0,000832679	0,4704039	-7,845025738
11442	0,000818708	0,482128741	-7,837326914
11397	0,00079772	0,49983564	-7,827228875
12203	0,001164337	0,226434094	-8,240905096
10309	0,000240401	0,888387469	-8,451547586
19	7,30849E-06	0,999862406	-11,82661112
10646	0,000425943	0,793551587	-7,992442716
10428	0,000308224	0,85975487	-8,235792266
11017	0,000614636	0,653906526	-7,819271397
11837	0,001000777	0,336498463	-7,996140429
13694	0,001796656	0,024739626	-10,02117701
11122	0,000665799	0,611600343	-7,806198765
12074	0,001107112	0,262183572	-8,144710414
10237	0,000198265	0,902388997	-8,628615461
11322	0,000762097	0,530059191	-7,81420289
10925	0,000568669	0,6907239	-7,842227188
11543	0,000865733	0,442921183	-7,866297055
11027	0,000619573	0,64987704	-7,8174521
10288	0,000228477	0,89264876	-8,497637317
11510	0,00085045	0,455578514	-7,85593192
10792	0,000501715	0,741493235	-7,89656845
10744	0,000476738	0,759357686	-7,923826057



	Ajusté	Initial
Lambda :	7,31E-06	1,00E-05
Bêta :	1,85315788	2
Eta :	1867,61618	2000
T :	9966,78908	10000

LN Vraisemblance : -850,829668

T	R(T) initial	N(t) initial	Fréquence	N(t) simulé	R(T) ajusté	N(t) ajusté
0	1	0	0	0	1	0
500	0,99501248	0,49875208	1	1	0,99635242	0,36475777

Figure 4. Simulation puis ajustement d'un modèle de Bertholon (100 valeurs)

Cet ajustement peut également s'effectuer à partir de données censurées en multipliant le produit des densités, pour les durées avant défaillance, par celui des fiabilités, pour les durées avant censure.

5 – Modèle de vieillissement à 3 phases

Le modèle de Bertholon 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. Il correspond alors à la mise en série de trois blocs, le premier régi par une loi de mortalité de Weibull, initiée à $t = 0$ ($\gamma = 0$) et limitée à une certaine durée, et les deux autres étant ceux du modèle de Bertholon. L'occurrence de la défaillance peut alors être simulée de la manière suivante :

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

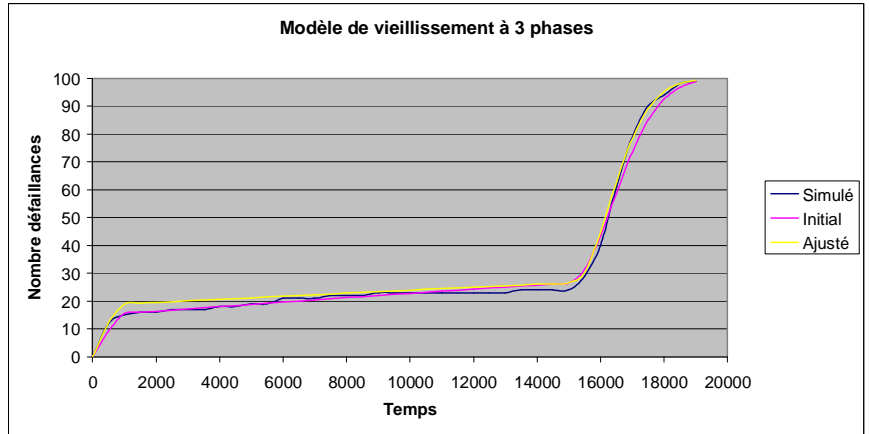
L'ajustement s'effectue de la même manière que précédemment avec pour expression de la densité :

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

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

A nouveau, on retrouve approximativement les valeurs des paramètres du modèle utilisé pour la simulation.

Valeurs simulées	$\lambda(t)$	R(t)	LN(f(t))
17879	0,001704401	0,059570533	-9,19513576
16602	0,000999617	0,337505248	-7,994312984
2087	6,9998E-06	0,804133105	-12,08762015
17318	0,001403475	0,142611235	-8,516436708
16125	0,000711457	0,508096721	-7,925279458
16607	0,001002714	0,335720031	-7,996522533
17864	0,00169631	0,061148021	-9,173757844
16905	0,001174247	0,242631291	-8,163340916
667	0,000165624	0,854848687	-8,862622444
41	0,000377753	0,978059026	-7,903453975
16136	0,000718686	0,503913251	-7,92343735
15801	0,000501821	0,618770747	-8,077286558
16981	0,001216923	0,221661673	-8,218032907
17049	0,00125511	0,203767923	-8,271305901
16258	0,000793898	0,459619173	-7,915912875
16931	0,00118841	0,235561762	-8,180921388
115	0,000277752	0,95533554	-8,234476162
32	0,000407678	0,981607206	-7,82359751
16382	0,000869291	0,41449236	-7,928533677
16935	0,001190697	0,234430178	-8,183813884
16634	0,001018478	0,326680225	-8,008219202
15707	0,000438571	0,646443389	-8,16825883
17217	0,00134828	0,16377001	-8,418218072
16422	0,000893024	0,400289051	-7,936465606
17027	0,001242582	0,209544912	-8,253380832
18528	0,0002039838	0,017651354	-10,23182777
16136	0,000718632	0,50394438	-7,923450286
5713	6,9998E-06	0,783983496	-12,11299699
16701	0,00105736	0,304742795	-8,04026689
17687	0,001602829	0,081818034	-8,939242833
16459	0,000915301	0,387002665	-7,945581457
16619	0,001009727	0,331689185	-8,001632593
3671	6,9998E-06	0,795271263	-12,0987017
15560	0,000333888	0,684503194	-8,383766233



	Ajusté	Initial
Bêta 1 :	6,95E-01	0,8
Eta 1 :	9999,9996	10000
T 1 :	1012,454212	1000
Lambda :	6,9998E-06	1,00E-05
Bêta 2 :	1,835146599	2
Eta 2 :	1638,152353	2000
Gamma 2 :	15184,7033	15000

LN Vraisemblance : -866,32265

T	R(T) initial	N(t) initial	Fréquence	N(t) simulé	R(T) ajusté	N(t) ajusté
0	1	0	0	0	1	0
500	0,908438389	9,156161148	12	12	0,879840131	12,0159869

6 – Conclusion

Un outil d'optimisation globale s'avère indispensable à l'ajustement d'un modèle paramétrique dès lors que celui-ci atteint une certaine complexité. Il rend alors 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. Il permet également d'effectuer des optimisations sur des paramètres divers tels que, par exemple, la durée d'amortissement des matériels ou l'occurrence optimale des actions de maintenance préventive.

Bibliographie

- [1] Fertou E., Procaccia H. (2008), Evaluation du temps d'initiation du vieillissement de matériels réparables et optimisation de la maintenance préventive, lambda mu 16, Avignon.
- [2] 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, Avignon.
- [3] A. Cabarbaye (2003), Outil générique d'optimisation dans le domaine discret et/ou continu éventuellement stochastique - ROADEF'03 – Avignon, 2003.
- [4] David E. Goldberg (1994), Algorithmes Génétiques, Exploration optimisation et apprentissage automatique, Addison-Wesley
- [5] V. Feoktistov, S. Janaqi (2004) - Evolution différentielle - Une vue d'ensemble, MOSIM 04, NANTES.