

EDF

Direction des Etudes et Recherches

**Electricité
de France**

SERVICE ENSEMBLES DE PRODUCTION
Département Retour d'Expériences Mesures-Essais
Département Surveillance Diagnostic Maintenance

Gestion INIS
Doc. enreg. le : 7/3/94
N° TRN : FR.95.00110
Destination : I.I+D.D

Novembre 1994

BACHA M.
CELEUX G.
IDEE E.
LANNOY A.
VASSEUR D.

**ESTIMATION BAYESIENNE DES PARAMETRES D'UNE
LOI DE WEIBULL**

**BAYESIAN ESTIMATION OF WEIBULL DISTRIBUTION
PARAMETERS**

VOL 27 No 17

Pages : 00011

95NB00110

Diffusion : J.-M. Lecœur
EDF-DER
Service IPN. Département SID
1, avenue du Général-de-Gaulle
92141 Clamart Cedex

© Copyright EDF 1995

ISSN 1161-0611

EDF

Direction des Etudes et Recherches

**Electricité
de France**

SERVICE ENSEMBLES DE PRODUCTION
Département Retour d'Expériences Mesures-Essais
Département Surveillance Diagnostic Maintenance

Gestion INIS

Doc. enreg. le : 7/3/94
N° TRN : FR 94000015
Destination : I.I+D.D

Novembre 1994

BACHA M.
CELEUX G.
IDEE E.
LANNOY A.
VASSEUR D.

**ESTIMATION BAYESIENNE DES PARAMETRES
D'UNE LOI DE WEIBULL**

***BAYESIAN ESTIMATION OF WEIBULL
DISTRIBUTION PARAMETERS***

Pages : 00011

95NB00110

Diffusion : J.-M. Lecœuvre
EDF-DER
Service IPN. Département SID
1, avenue du Général-de-Gaulle
92141 Clamart Cedex

© Copyright EDF 1995

ISSN 1161-0611

SYNTHÈSE :

Dans cette note, sont exposées les méthodes SEM (Stochastic Expectation Maximization) et WLB-SIR (Weighted Likelihood Bootstrap - Sampling Importance Resampling), pour l'estimation des paramètres de forme et d'échelle d'une loi de Weibull très censurée.

La seconde méthode est bayésienne et permet ainsi de tenir compte des informations a priori disponibles sur les paramètres à estimer.

Une application sur un jeu de données réelles provenant du retour d'expérience des centrales nucléaires à eau sous pression a été réalisée.

BACHA M., CELEUX G. : INRIA
IDEE E. : Université de Savoie
LANNOY A. : EDF-DER, Département SDM
VASSEUR D. : EDF-DER, Département REME

EXECUTIVE SUMMARY :

In this paper, we expose SEM (Stochastic Expectation Maximization) and WLB-SIR (Weighted Likelihood Bootstrap - Sampling Importance Resampling) methods which are used to estimate Weibull distribution parameters when data are very censored.

The second method is based on bayesian inference and allow to take into account available prior informations on parameters.

An application of this method, with real data provided by nuclear power plants operation feedback analysis has been realized.

SOMMAIRE

1. INTRODUCTION	4
2. RAPPEL DE QUELQUES METHODES CLASSIQUES	5
3. ESTIMATION BAYESIENNE.....	6
3.1. DES INFORMATIONS A PRIORI AUX LOIS A PRIORI.....	6
3.2. ALGORITHME WLB - SIR.....	7
4. RESULTATS NUMERIQUES.....	7
4.1. DONNEES SIMULEES	7
4.2. DONNEES REELLES	8
5. CONCLUSION	8
REFERENCES.....	9

ESTIMATION BAYESIENNE DES PARAMETRES D'UNE LOI DE WEIBULL

M. BACHA, G. CELEUX, E. IDEE, A. LANNOY & D. VASSEUR

1. INTRODUCTION

On considère un équipement industriel dont la durée de vie suit une loi de Weibull de paramètre de forme β et de paramètre d'échelle η , $W(t|\theta = (\beta, \eta))$ de densité

$$f(t|\theta) = R(t|\theta) \cdot h(t|\theta), \beta > 0, \eta > 0 \quad (1)$$

où

$$R(t|\theta) = \exp\left(-\left(\frac{t}{\eta}\right)^\beta\right) \quad (2)$$

et

$$h(t|\theta) = \frac{\beta}{\eta} \left(\frac{t}{\eta}\right)^{\beta-1} \quad (3)$$

sont respectivement la fonction de survie et le taux de défaillance instantané.

Pour tout calcul des paramètres caractéristiques de sûreté de fonctionnement telles que disponibilité, fiabilité, taux de défaillance ..., il est nécessaire de connaître les paramètres β et η ou du moins d'en avoir de bonnes approximations. C'est à ce problème d'estimation des paramètres qu'on va donner une réponse tout d'abord fréquentielle puis bayésienne.

Soit $t = (t_1, \dots, t_r, t_{r+1}, \dots, t_n)$ un échantillon de la loi de probabilité $W(t|\theta)$ qui modélise la durée de vie du matériel étudié. Les observations t_1, \dots, t_r désignent des dates de panne (perte de fonction complète du matériel étudié) et t_{r+1}, \dots, t_n des dates de censure généralement différentes (opérations de maintenance préventive, fin d'observation de l'échantillon). Souvent n est relativement petit, pour des raisons économiques : coût des essais de fiabilité, pour des raisons de qualité des matériels : matériels très fiables. Par contre, le nombre $(n-r)$ de censures peut être grand car les matériels ont une moyenne de vie assez grande. On dispose quelquefois d'informations sur les paramètres (β, η) soit par avis d'experts soit par le retour d'expérience passé des matériels similaires. Il est donc utile de pouvoir inclure ces informations complémentaires dans la méthode d'estimation utilisée afin d'améliorer la qualité des estimateurs.

Dans la suite, nous allons décrire brièvement les méthodes couramment utilisées en estimation (cf. [2]). Puis nous détaillerons une méthode utilisant l'algorithme WLB-SIR qui permet de tenir compte des informations disponibles a priori. Ces méthodes seront illustrées par des exemples de retour d'expérience de matériels de centrales nucléaires à eau sous pression.

2. RAPPEL DE QUELQUES METHODES CLASSIQUES

La méthode la plus utilisée est la méthode de maximum de vraisemblance. Elle consiste à estimer $\theta^* = (\beta^*, \eta^*)$, le vrai paramètre, par $\hat{\theta} = (\hat{\beta}, \hat{\eta})$, argument qui réalise le maximum de vraisemblance

$$L(\theta; t) = \prod_{i=1}^r f(t_i | \theta) \prod_{i=r+1}^n R(t_i | \theta) \quad (4)$$

L'estimateur $\hat{\theta} = (\hat{\beta}, \hat{\eta})$ est donné par le système d'équations dont la première se résout itérativement

$$\frac{1}{\hat{\beta}} \cdot \frac{\sum_{i=1}^r \log(t_i)}{r} - \frac{\sum_{i=1}^n t_i^{\hat{\beta}} \log(t_i)}{\sum_{i=1}^n t_i^{\hat{\beta}}} = 0 \quad (5)$$

et

$$\hat{\eta} = \left[\frac{\sum_{i=1}^n t_i^{\hat{\beta}}}{r} \right]^{1/\hat{\beta}} \quad (6)$$

Les variances asymptotiques peuvent être approximées par (cf. [4])

$$\sigma_{\hat{\beta}}^2 = \frac{0.608}{r} \hat{\beta}^2 \quad (7)$$

et

$$\sigma_{\hat{\eta}}^2 = \frac{1.087}{r} \left(\frac{\hat{\eta}}{\hat{\beta}} \right)^2 \quad (8)$$

On remarque que ces variances sont inversement proportionnelles au nombre r des observations non censurées. D'autre part, il est connu que les estimateurs du maximum de vraisemblance ne sont précis que pour n et r tendant vers l'infini (cf. [6]). L'application de cette méthode peut donc s'avérer décevante dans certains cas où r (nombre de défaillances) est petit. Une méthode plus appropriée est la méthode SEM (Stochastic Expectation Maximization).

L'algorithme SEM permet d'estimer les paramètres à partir d'un échantillon incomplet et ce de façon plus sûre que celle de la méthode du maximum de vraisemblance. L'idée de base de la méthode est simple : au lieu de travailler avec l'échantillon incomplet réellement observé, on travaille avec des échantillons, duplicatas de l'échantillon de départ, dans lesquels on a remplacé les données manquantes par des données simulées. L'estimation résultante sera fonction des différents estimateurs du maximum de vraisemblance, calculés pour ces échantillons complétés aléatoirement. Formellement, l'algorithme SEM appliqué à un échantillon censuré, consiste à

- initialiser avec $\theta^{(0)} = (\beta^{(0)}, \eta^{(0)})$,

et à l'itération ($i = 1$)

- compléter l'échantillon observé en remplaçant toutes les dates de censure par des simulations de valeurs plus grandes que la censure correspondante suivant la loi $W(t | \beta^{(i)}, \eta^{(i)})$,

- calculer $\beta^{(i+1)}$ et $\eta^{(i+1)}$ qui sont les estimateurs du maximum de vraisemblance de l'échantillon ainsi complété.

Après M itérations, on obtient $\theta^{(1)}, \dots, \theta^{(M)}$ réalisations d'une chaîne de Markov homogène et ergodique, qui, pour M grand, s'approche de son unique distribution stationnaire. Sur le plan pratique, le fait que la chaîne soit ergodique implique les deux choses importantes suivantes :

1. Le paramètre $\theta^{(0)}$ d'initialisation peut être quelconque. Toutefois, il est conseillé de choisir une valeur raisonnable (pas trop éloignée de la vraie valeur) pour que la stationnarité ait lieu vite, sinon il faut prendre M vraiment grand pour être assuré d'avoir atteint le régime stationnaire.
2. Les valeurs moyennées suivantes :

$$\beta_{sem} = \frac{1}{L} \sum_{i=M-i+1}^M \beta^{(i)} \quad (9)$$

et

$$\eta_{sem} = \frac{1}{L} \sum_{i=M-i+1}^M \eta^{(i)} \quad (10)$$

pour $L > M$ assez grands, sont de bonnes estimations des paramètres.

Pour plus de détails concernant cette méthode, on pourra consulter [2] et [3].

3. ESTIMATION BAYESIENNE

Toute estimation bayésienne se compose de deux parties distinctes. La première, subjective, est l'étape de modélisation des informations a priori par des lois de probabilité adéquates ; elle fait appel beaucoup plus à l'expérience et au bon sens du statisticien qu'à des considérations mathématiques. La seconde étape, objective, est l'étude des lois a posteriori pour en tirer les estimateurs désirés.

3.1. Des informations a priori aux lois a priori

Dans cette partie, nous supposons que l'expert est capable de fournir les informations suivantes :

- un intervalle d'appartenance $[\beta_g, \beta_d]$ pour le paramètre de forme β ,
- une approximation β_0 du vrai paramètre de forme β^* ,
- une approximation du vrai paramètre d'échelle η^* , soit η_0 .

Pour le paramètre d'échelle η , nous prenons comme loi a priori $\pi(\eta)$ la loi gamma $\mathcal{G}(a,b)$ de densité

$$g(\eta) = \frac{\eta^{a-1} \exp(-(\eta/b))}{\Gamma(a) b^a}, \quad \eta \geq 0$$

dont la moyenne, la variance et le mode valent respectivement a , ab^2 et $(a-1)b$ pour $a \geq 1$.

Les hyperparamètres a et b sont donnés par les relations $a b^2 = \frac{E \beta_0^2}{r \Gamma_0 K}$ et $a b = \eta_0$ (resp. $(a-1)b = \eta_0$) si

η_0 est la valeur moyenne (resp. la valeur a priori la plus probable) de η . Pour le paramètre de forme β , on prend comme loi a priori $\pi(\beta)$ une loi gamma $\mathcal{G}(c,d)$ où les hyperparamètres c et d sont choisis tels que cette loi soit concentrée avec une forte probabilité (0.99 pour notre étude numérique) sur l'intervalle $[\beta_g, \beta_d]$. A cette fin, on utilise l'approximation, valable pour c suffisamment grand, de la loi gamma par une loi normale.

Les lois a posteriori sont alors :

$$\pi(\beta|t) \propto L(\theta;t) \pi(\beta)$$

et

$$\pi(\eta|t) \propto L(\theta;t) \pi(\eta)$$

Les caractéristiques de ces lois (moyenne, variance, ...) ne peuvent être calculées explicitement et le recours à des intégrations avec simulation des paramètres β et η est inévitable.

3.2. Algorithme WLB - SIR

L'algorithme WLB-SIR (Weighted Likelihood Bootstrap - Sampling Importance Resampling, cf. [7]) permet de simuler de façon approximative des lois d'expressions complexes et ce d'une manière non itérative par opposition aux méthodes de type MCMC (Markov Chain Monte Carlo, cf. [5]) tel que le Gibbs sampling.

Pour expliquer le fonctionnement de l'algorithme, notons $\pi(\theta|t) \propto L(\theta;t)\pi(\theta)$ la loi a posteriori qu'on cherche à simuler, $\pi(\theta)$ et $L(\theta;t) = \prod_{i=1}^r f(t_i|\theta) \prod_{i=r+1}^n R(t_i|\theta)$ désignant respectivement la loi a priori du

paramètre θ et la vraisemblance de l'échantillon observé. L'étape WLB consiste à faire

- fixer un réel positif α ,
- simuler un échantillon de poids $w_1^{*i} = (w_1, \dots, w_n)$ suivant la loi de probabilité de Dirichlet uniforme,
- calculer $\hat{\theta}$ argument du maximum de $\hat{E}(\theta) = \prod_{i=1}^r [f(t_i|\theta)^{w_i}]^\alpha \cdot \prod_{i=r+1}^n [R(t_i|\theta)^{w_i}]^\alpha$

En répétant M fois cette étape, on obtient l'échantillon $\hat{\theta}_1, \dots, \hat{\theta}_M$. Par la méthode non paramétrique des noyaux

(cf. (8)), on peut estimer la densité de la loi de probabilité dont il est issu, soit \hat{g} cette densité estimée. L'étape SIR dont le but est de rapprocher \hat{g} de la loi a posteriori $\pi(\theta|t)$ se résume à :

- pour $i = 1, \dots, M$, calculer les poids $p_i \propto \pi(\hat{\theta}_i)L(\hat{\theta}_i;t) / \hat{g}(\hat{\theta}_i)$,
- tirer $(\theta_1, \dots, \theta_m)$ avec les probabilités (p_1, \dots, p_M) de l'échantillon $(\hat{\theta}_1, \dots, \hat{\theta}_M)$.

Quand le rapport (M/m) devient suffisamment grand, l'échantillon $(\theta_1, \dots, \theta_m)$ suit la loi a posteriori $\pi(\theta|t)$.

Il incombe à l'utilisateur de choisir le paramètre α . Par défaut ce paramètre est fixé à 1. Notons qu'un $\alpha > 1$ donne une distribution dispersée des w_i^* , alors que $\alpha < 1$ rend cette distribution concentrée (cf. [1]). De plus, ce n'est pas une obligation d'utiliser la loi de Dirichlet dans l'étape WLB, d'autres distributions peuvent être utilisées pour le tirage des poids.

4. RESULTATS NUMERIQUES

Dans cette partie, nous traitons deux échantillons, le premier simulé et le second réel concernant un matériel d'une centrale nucléaire à eau sous pression.

4.1. Données simulées

Nous considérons un échantillon de taille 25 de la loi de Weibull de forme $\beta^* = 1.2$ et d'échelle $\eta^* = 100$, censuré à la date fixe $c = 40$. Le nombre r d'observations non censurées est 4 (ces observations sont : 2.91 ; 6.89 ; 12.41 et 25.47). Les estimateurs de maximum de vraisemblance sont $\hat{\beta} = 0.703$ et $\hat{\eta} = 469.503$. Les informations a priori disponibles sont β^* , qui est supposé appartenir à l'intervalle $(0.5, 3.0]$ et approximé par $\beta_0 = 1.5$ et η^* est approximé par $\eta_0 = 120$. En calculant les moyennes empiriques des 100 dernières itérations sur 300 générées par l'algorithme SEM initialisé avec $\theta^{(0)} = (\beta_0, \eta_0)$, on a $\beta_{sem} = 1.43$ et $\eta_{sem} = 167.28$. Pour les calculs bayésiens, les lois a priori sont $\pi(\eta) = \mathcal{G}(9.0, 13.3)$ et $\pi(\beta) = \mathcal{G}(18.0, 0.1)$. Pour trois choix différents du paramètre α , l'étape WLB a été exécutée 5000 fois puis l'étape SIR 1000 fois. Les moyennes et écarts types des lois a posteriori sont donnés dans la table n° 1 et les densités des lois a priori et a posteriori de β et η pour $\alpha = 1$ sont tracés sur la figure n° 1 (voir dernière page).

α	E(β)	Std(β)	E(η)	Std(η)
0.6	1.06	0.16	170.56	31.80
1.0	1.18	0.21	142.21	31.52
1.4	1.27	0.24	132.49	31.67

Table 1. Données simulées, moyennes (E) et écarts type (Std) des lois a posteriori pour différentes valeurs de α

4.2. Données réelles

Le deuxième jeu de données concerne un matériel provenant d'une centrale nucléaire à eau sous pression. Il s'agit du dispositif de chargement en air de lancement du dispositif de démarrage des diesels de secours. C'est un échantillon de taille 248 dont seulement 136 observations ne sont pas censurées. Les dates de censure correspondent soit à des dégradations (défaillances partielles), soit à des fins d'observation. Avec les mêmes notations qu'avant et en appliquant la même démarche que pour les données simulées, on trouve $\hat{\beta} = 2.31$ et $\hat{\eta} = 66321.60$. On a pris $\beta_0 = 2.0$ et $\eta_0 = 70000$ avec l'hypothèse que β^* est dans l'intervalle $[1.0, 3.0]$. Les estimateurs donnés par la méthode SEM sont $\beta_{sem} = 2.31$ et $\eta_{sem} = 66181.56$. Les lois a priori considérées sont $\pi(\beta) = \mathcal{G}(36, 0.06)$ et $\pi(\eta) = \mathcal{G}(511.00, 128.68)$. Sur la table n° 2, on a reporté les moyennes et écarts type empiriques des lois a posteriori. La figure n° 2 (voir dernière page) représente les densités a priori et a posteriori des paramètres β et η pour $\alpha = 1$.

α	E(β)	Std(β)	E(η)	Std(η)
0.6	2.29	0.07	66897.28	1216.84
1.0	2.27	0.10	67357.66	1500.72
1.4	2.28	0.11	67492.16	1799.46

Table 2. Données réelles, moyennes (E) et écarts type (Std) des lois a posteriori pour différentes valeurs de α

On remarque que les résultats sont similaires, quelle que soit la méthode utilisée, ce qui s'explique par la grande quantité d'informations amenées par le retour d'expérience (défaillances nombreuses constatées).

5. CONCLUSION

Les simulations montrent que les estimateurs du maximum de vraisemblance peuvent être très décevants pour de petits échantillons très censurés. La méthode bayésienne est plus performante que les méthodes classiques d'estimation dans le cas d'échantillons de faible taille ou excessivement censurés. Toutefois, l'étape de modélisation de l'information a priori doit être traitée avec précaution, une information a priori erronée ou une modélisation inadéquate de cette information risque de donner de mauvais résultats. Le choix du paramètre α (éventuellement celui de la distribution des poids) dans la méthode WLB-SIR, peut être crucial. Aussi, l'algorithme SEM, qui fournit ici des estimations proches des estimations bayésiennes peut être envisagé comme une méthode intéressante lorsque l'information a priori est inexistante ou sujette à caution.

REFERENCES

- [1] BACHA M. & CELEUX G. (1994). Contribution à la discussion de papier de Newton, M.A. & Raftery, A.E., *J. Roy. Statist. Soc., Ser. B*, Vol. 56, Num. 2.
- [2] BCHA M., CELEUX G., IDEE E., LANNON A., & VASSEUR D (1992). Détermination de la loi de dégradation d'un matériel et de ses composants pour l'aide à l'augmentation de la maintenance préventive. Contrat P16/4M0277/E8537 - EDF-INRIA.
- [3] CELEUX G, & DIEBOLT J. (1985). The SEM algorithm: a probabilistic teacher algorithm derived from the EM algorithm for the mixture problem. *Computational Statistics Quarterly*, Vol. 2.
- [4] COHEN A.C. (1988). Parameter estimation in reliability and life span models. *Statistics: textbooks and monographs*, Ed. Dekker, M.
- [5] GELMAN A., GEYER C.J. & RUBIN D.B. (1992). Iterative simulation using single and multiple sequences. *Statistical Science*, Vol. 7, Num. 4.
- [6] LAWLESS J.F. (1982). *Statistical models and methods for lifetime data*. Wiley, New York.
- [7] NEWTON M.A. & RAFTERY A.E. (1994). Approximate Bayesian inference with the weighted likelihood bootstrap (avec discussion). *J. Roy. Statist. Soc., Ser. B*, Vol. 56, Num. 1.
- [8] TERREL G.R. (1990). The maximal smoothing principle in density estimation, *J. Am. Statist. Ass.*, Num. 85.

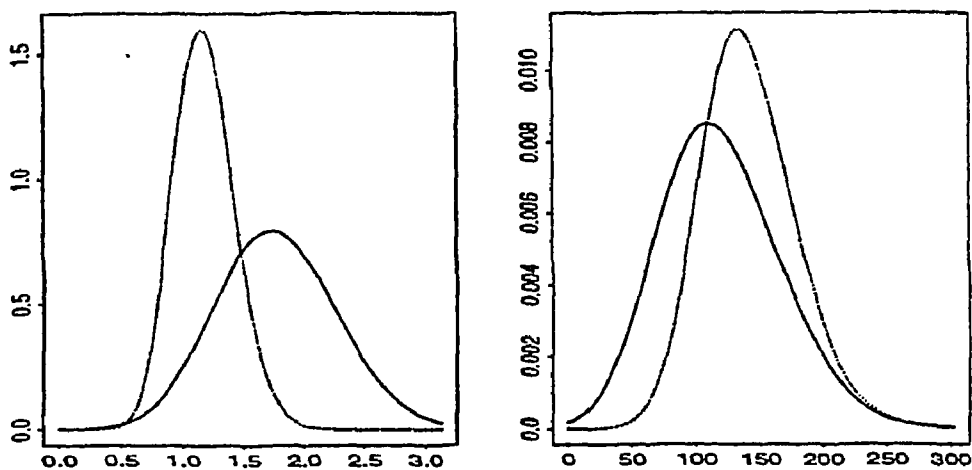


Fig. 1. Données simulées, densités des lois a priori (ligne continue) et a posteriori (en pointillé) des paramètres β (à gauche) et η (à droite) pour $\alpha = 1$.

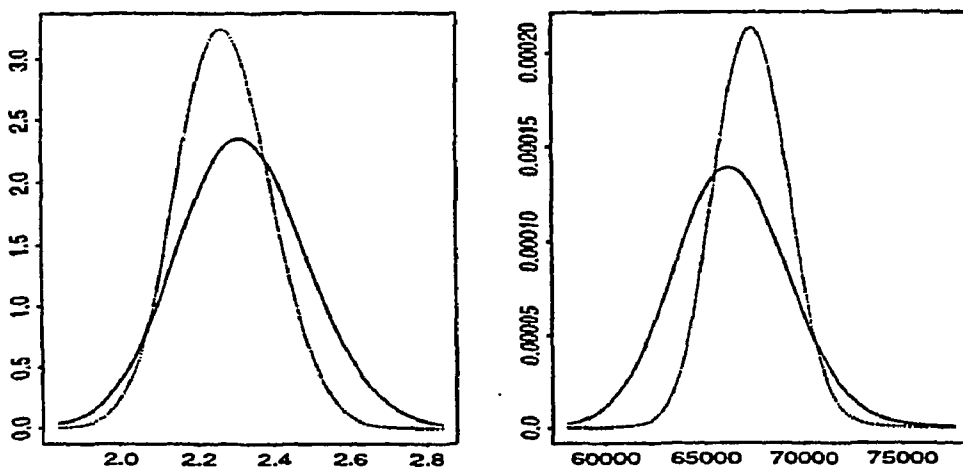
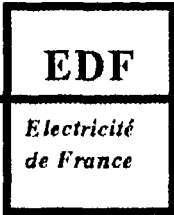


Fig. 2. Données réelles, densités des lois a priori (ligne continue) et a posteriori (en pointillé) des paramètres β (à gauche) et η (à droite) pour $\alpha = 1$.



Direction des Etudes
et Recherches

Service Information
Prospective et Normalisation

CLAMART Le 17/01/96

Département Systèmes d'information
et de documentation

CEA
MIST/SBDS/SPRI
CENTRE DE SACLAY

Groupe Exploitation
de la Documentation Automatisée

1, avenue du Gal de Gaulle
92141 CLAMART Cedex
tel : 47 65 56 33

91191 GIF SUR YVETTE CEDEX

à l'attention de :

MEMOIRE TECHNIQUE ELECTRONIQUE

Cette feuille est détachable grâce à la microperforation sur le coté droit.

Référence de la demande : **F558899**
Origine : **CATALOGUE DES NOTES DER**

Votre commande :

Numéro du document : **95NB00110**

Titre : **ESTIMATION BAYESIENNE DES PARAMETRES D'UNE LOI DE WEIBULL**

Auteurs : **BACHA M./CELEUX G./IDEE E./LANNOY A./VASSEUR D.**

Source : **COLL. NOTES INTERNES DER. PRODUCTION D'ENERGIE (HYDRAULIQUE, THE**
Serial :

Référence du document : **SANS**

Nombre de pages: **0012**

Nombre d'exemplaires : **001**

Support : **P**