



# Comparaison de méthodes d'ajustement d'une distribution de Weibull à 3 paramètres sur une base de données de mesures de ténacité

**M. Marquès, N. Pérot, N. Devictor**

Laboratoire de Conduite et Fiabilité des Réacteurs

DEN/DER/SESI

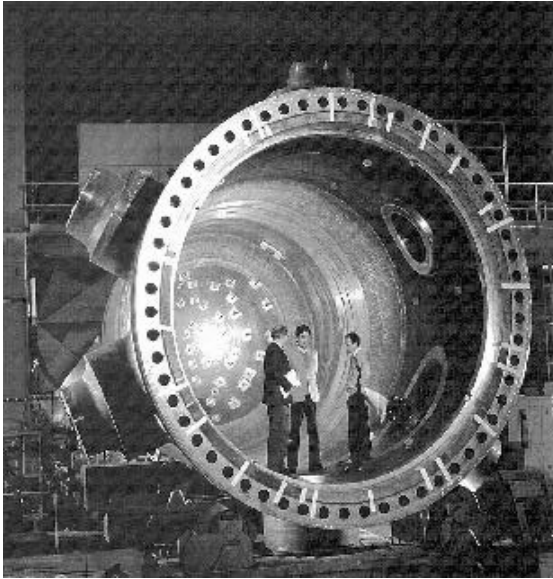
Centre d'Etude Nucléaire de Cadarache

13108 St-Paul-lez-Durance Cedex

---

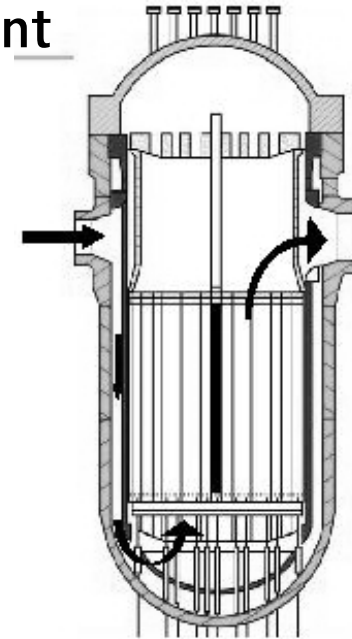
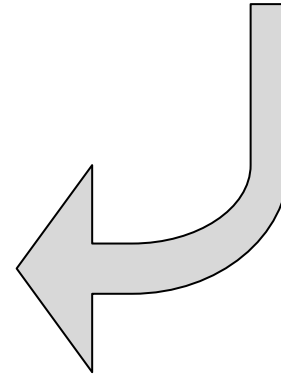
*CADARACHE*

# Contexte et Objectifs



The knowledge of the behaviour of the pressure vessel is of prime importance in the nuclear power plant lifetime program

Evaluation of the failure risk for a reactor vessel subjected to pressurised thermal shock transient



The scatter in the results of steel toughness measures is modelised (Master curve) by a Weibull distribution

$$F_{K_{IC}}(K) = P(K_{IC} \leq K) = 1 - \exp\left(-\left(\frac{K - K_{\min}}{K_0 - K_{\min}}\right)^\alpha\right)$$

# Base de données de ténacité EPRI

---

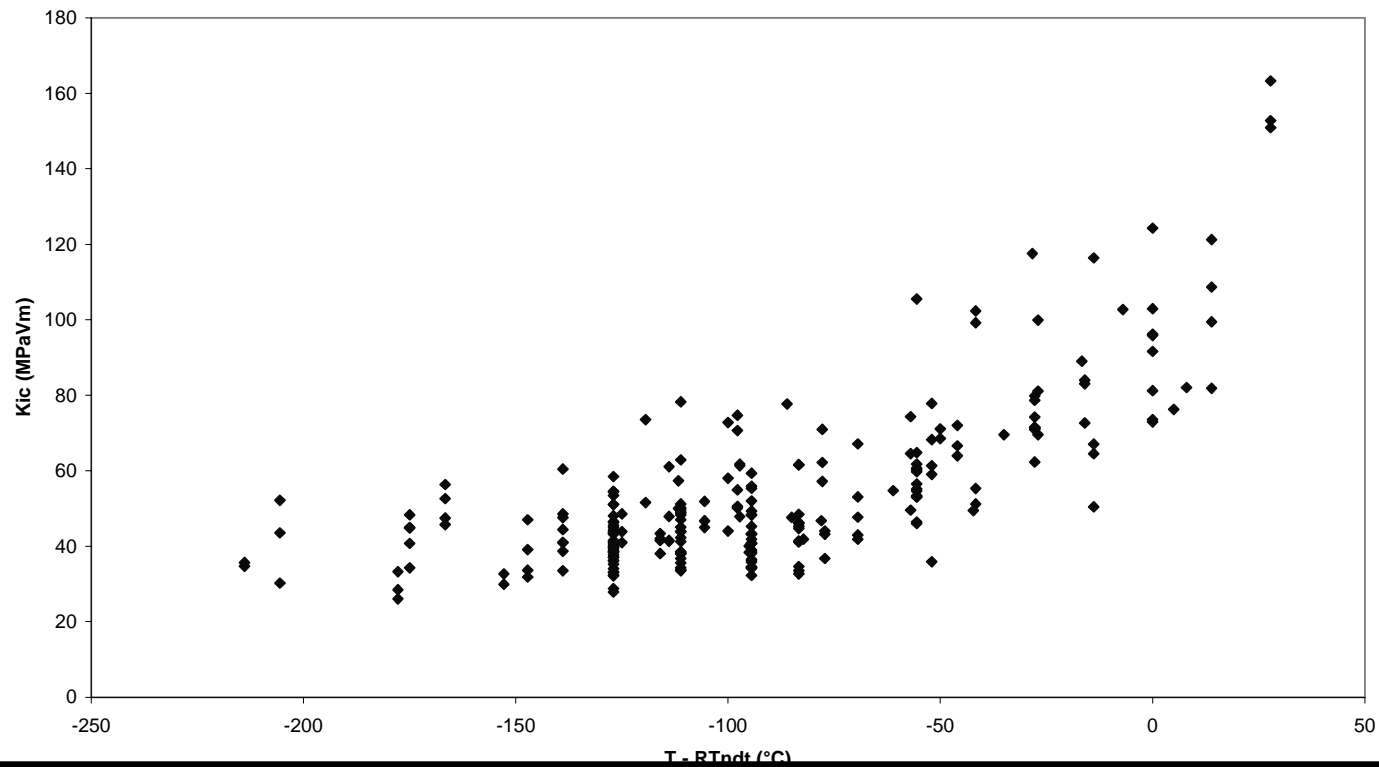
Base de données EPRI :



254 données de ténacité  $K_{Jc}$ , 18 matériaux,

$26 < K_{Jc} < 163$  MPa

Normalisés par  $RT_{NDT}$ :  $-214^{\circ}\text{C} < T - RT_{NDT} < 18^{\circ}\text{C}$



## Context and Objectives

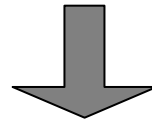
---



Characterisation of the toughness databases  
by the master Curve model does not seem sufficient

$\alpha = 4$ ,  $K_{min} = 20 \text{ MPa}\sqrt{m}$  *too conservative*

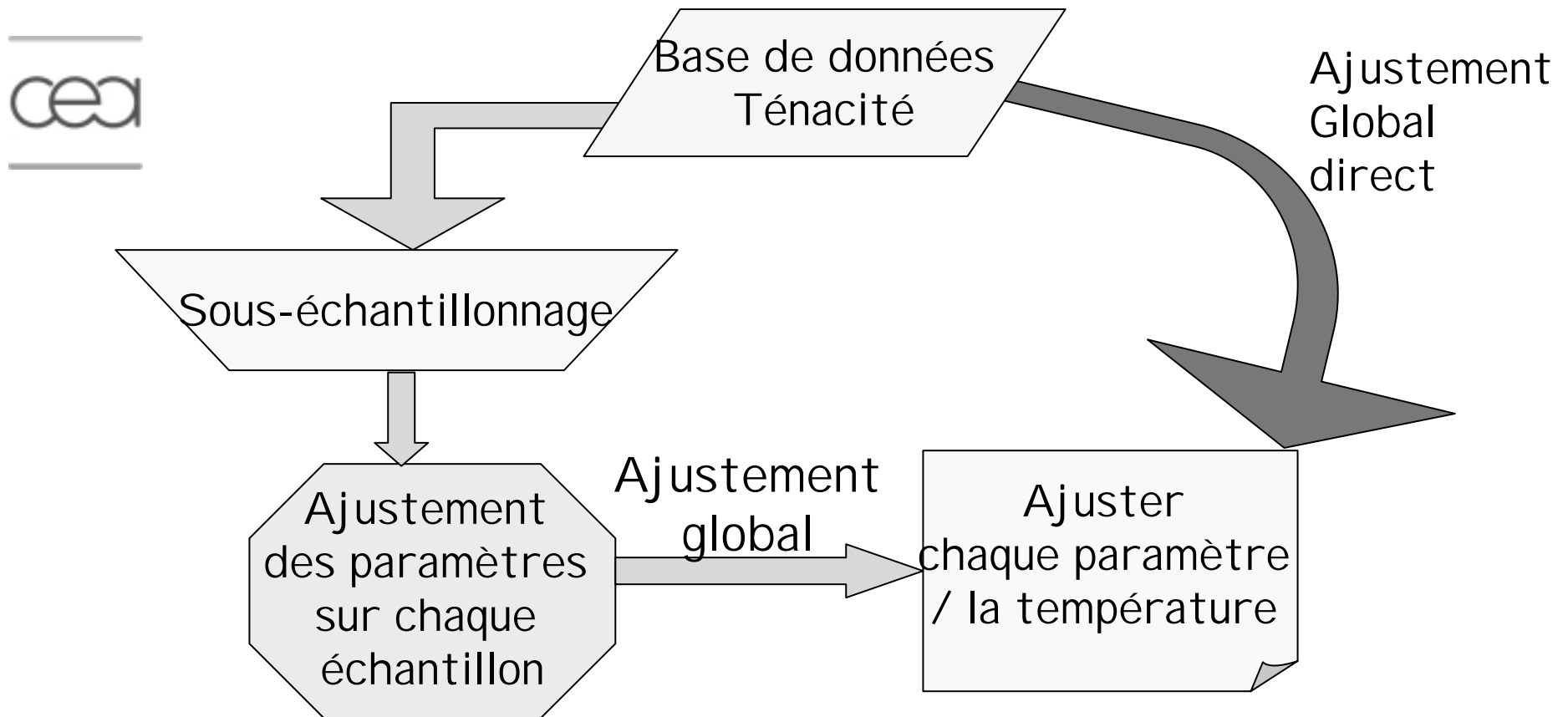
Toughness is very dependent on material temperature



Developpement of a methodology  
based on 5 estimation techniques for fitting a  
3-parameter Weibull distribution  
on a toughness database

The result of these methods is a Weibull  
distribution where the 3 parameters are  
functions of the temperature

# Description de la méthodologie

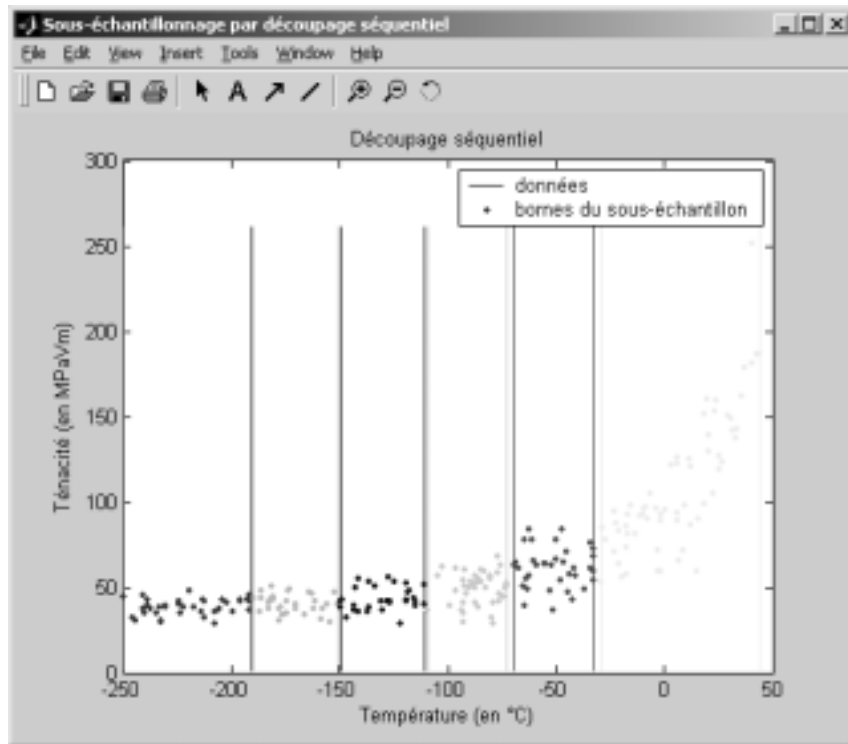


$$F_{K_{IC}}(K) = 1 - \exp\left(-\left(\frac{K - \mu(T)}{\beta(T)}\right)^{\alpha(T)}\right)$$

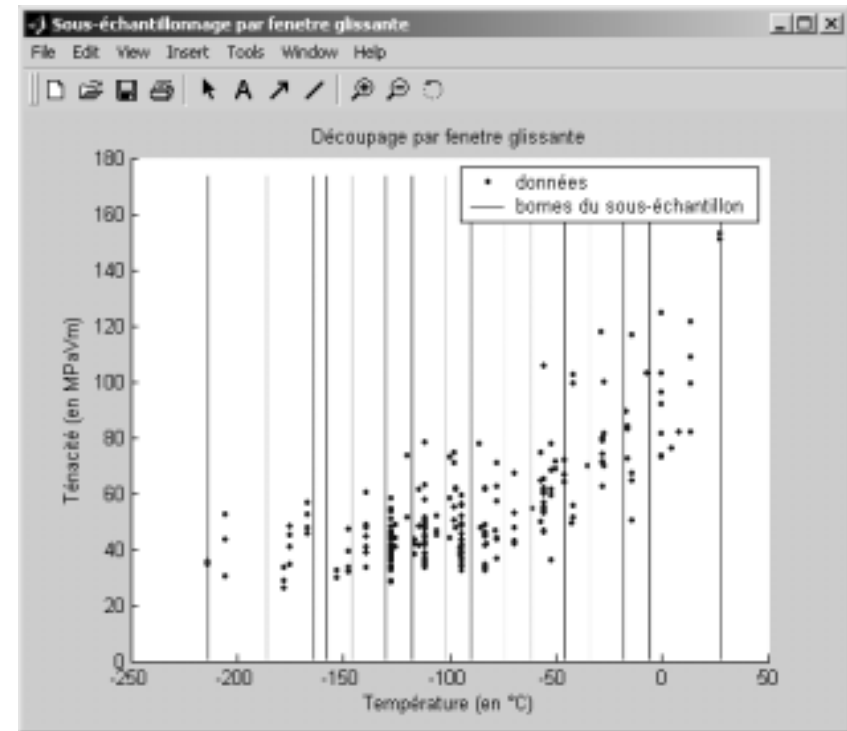
# Méthodes d'ajustement nécessitant un sous-échantillonnage



Base de données de ténacité  
US Electric Power  
Research Institute (EPRI)



Sous-échantillonnage séquentiel



Sous-échantillonnage par fenêtre glissante

# Méthodes d'ajustement local sur chaque échantillon



1. Ajustement basé sur la méthode des moments
2. Ajustement basé sur la méthode du maximum de vraisemblance
3. Ajustement basé sur le mixage des méthodes 1 et 2
4. Ajustement basé sur le maximum de vraisemblance estimé par un algorithme génétique

# Ajustement basé sur la méthode des moments

---



1. Pour  $\alpha$  (paramètre de forme) variant sur  $[a;b]$  avec un pas  $h$ , l'estimation de  $K_0-K_{min}$  et  $K_{min}$  est obtenue par la méthode des moments.
2. Seuls les triplets  $(\alpha, (K_0-K_{min})^*, K_{min}^*)$  ayant un sens physique sont acceptés.
3. Un critère de qualité de l'ajustement est calculé :
  - Cramer von Mises (CMV),
  - Kolmogorov-Smirnov (K-S)
  - Aderson-Darling (AD)
4. Le triplet  $(\alpha, (K_0-K_{min})^*, K_{min}^*)$  sélectionné minimise le critère sélectionné.



## Ajustement basé sur la méthode du maximum de vraisemblance

---



Fonction de maximum de vraisemblance pour la distribution de Weibull avec  $(x_1, \dots, x_n)$  l'échantillon

$$L(\alpha, K_0 - K_{\min}, K_{\min}) = \prod_{i=1}^n \left\{ \frac{\alpha}{K_0 - K_{\min}} \left( \frac{x_i - K_{\min}}{K_0 - K_{\min}} \right)^{\alpha-1} \exp \left[ - \left( \frac{x_i - K_{\min}}{K_0 - K_{\min}} \right)^{\alpha} \right] \right\}$$

1. Pour tout  $K_{\min}$  (paramètre de position) variant sur  $[a;b]$  avec un pas  $h$ , l'estimation de  $\alpha$  et  $K_0 - K_{\min}$  est obtenu par la méthode du maximum de vraisemblance.
2. *Un critère de qualité de l'ajustement est calculé :*
  - *Cramer von Mises (CMV)*
  - *Kolmogorov-Smirnov (K-S)*
  - *Aderson-Darling (AD)*
3. *Le triplet  $(\alpha^*, (K_0 - K_{\min})^*, K_{\min}^*)$  sélectionné minimise le critère choisi.*

# Ajustement basé sur une méthode combinée MM-ML

---

Williams and Dickson (1982)

1. Initialisation de la valeur de  $K_{min}$  (paramètre de position). Cette valeur doit être inférieure à la valeur minimum observée dans l'échantillon.
2. Pour cette valeur  $K_{min}$ , Les estimations du paramètre d'échelle  $(K_0 - K_{min})^*$  et du paramètre de forme  $\alpha^*$  sont obtenus par ML.
3. Pour l'estimateur ML  $\alpha^*$ , calcul par la méthode MM du paramètre de position  $K_{min}^{**}$  et du paramètre d'échelle  $(K_0 - K_{min})^{**}$ .

4. Soit

$$\delta = \left| \frac{K_{min} - K_{min}^{**}}{K_{min}} \right|$$

Tant que  $\delta > \varepsilon$  (tolerance), la position du paramètre  $K_{min}$  est incrémentée par une valeur fixée  $h$  et les étapes 1 à 3 sont répétées.

5. Si  $\delta \leq \varepsilon$ , le triplet  $(\alpha^*, (K_0 - K_{min})^*, K_{min}^{**})$  est sélectionné comme meilleure estimation des 3 paramètres de l'échantillon.
- 



# Ajustement basé sur l'estimation de ML par un algorithme génétique

---



Cette méthode consiste à calculer  
La meilleure estimation  $(\alpha^*, (K_0 - K_{min})^*, K_{min}^*)$   
des 3 paramètres  
en maximisant directement le logarithme de  
la fonction vraisemblance  
Avec un algorithme génétique  
sur chaque échantillon

# Description de la méthode avec algorithme génétique

Critère à maximiser : Adaptation = log-vraisemblance



Soient les  $n$  données  $\{X; T\} = \{x_i; t_i\}_{i=1, \dots, n}$

$$\mu_i = g(t_i; a_1, b_1, c_1)$$

$$\alpha_i = h(t_i; a_2, b_2, c_2)$$

$$\gamma_i = k(t_i; a_3, b_3, c_3)$$

Log-vraisemblance  $\ln L(X; \mu, \alpha, \gamma, T) = \sum_i^n \ln f(x_i; \mu_i, \alpha_i, \gamma_i)$

$$f(x_i; \mu_i, \alpha_i, \gamma_i) = \frac{\alpha_i}{\gamma_i} \left( \frac{x_i - \mu_i}{\gamma_i} \right)^{\alpha_i - 1} \exp \left( - \left( \frac{x_i - \mu_i}{\gamma_i} \right)^{\alpha_i} \right)$$

# Description de la méthode avec algorithme génétique

A chaque génération



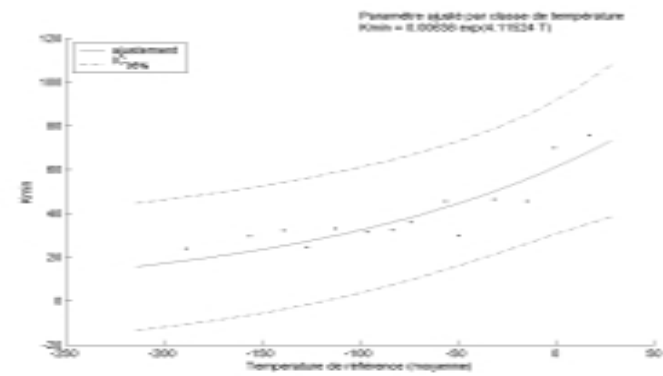
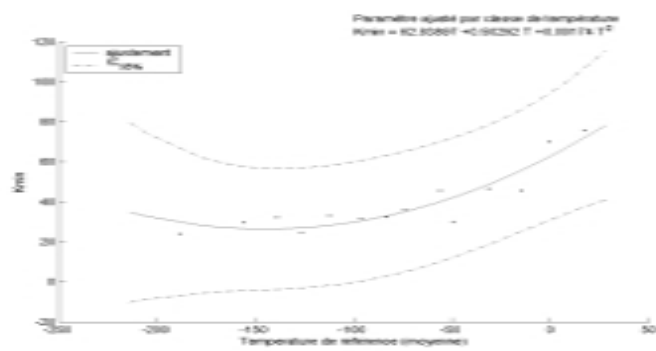
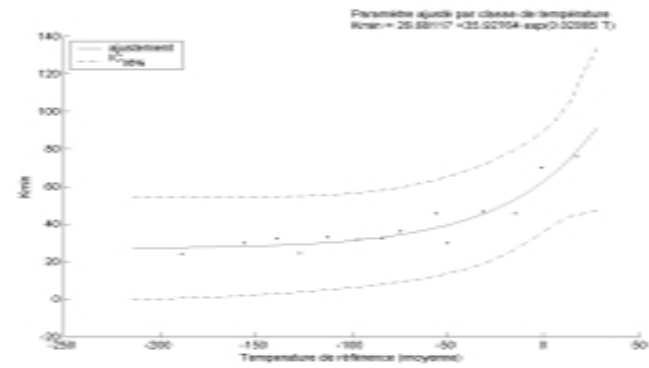
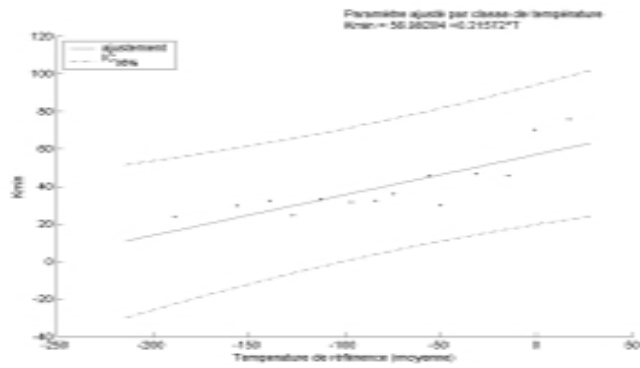
**Sélection** : désigne les plus aptes à survivre et à transmettre leurs gènes à la génération suivante

**Croisement** : échange de gènes au hasard pour produire 2 descendants

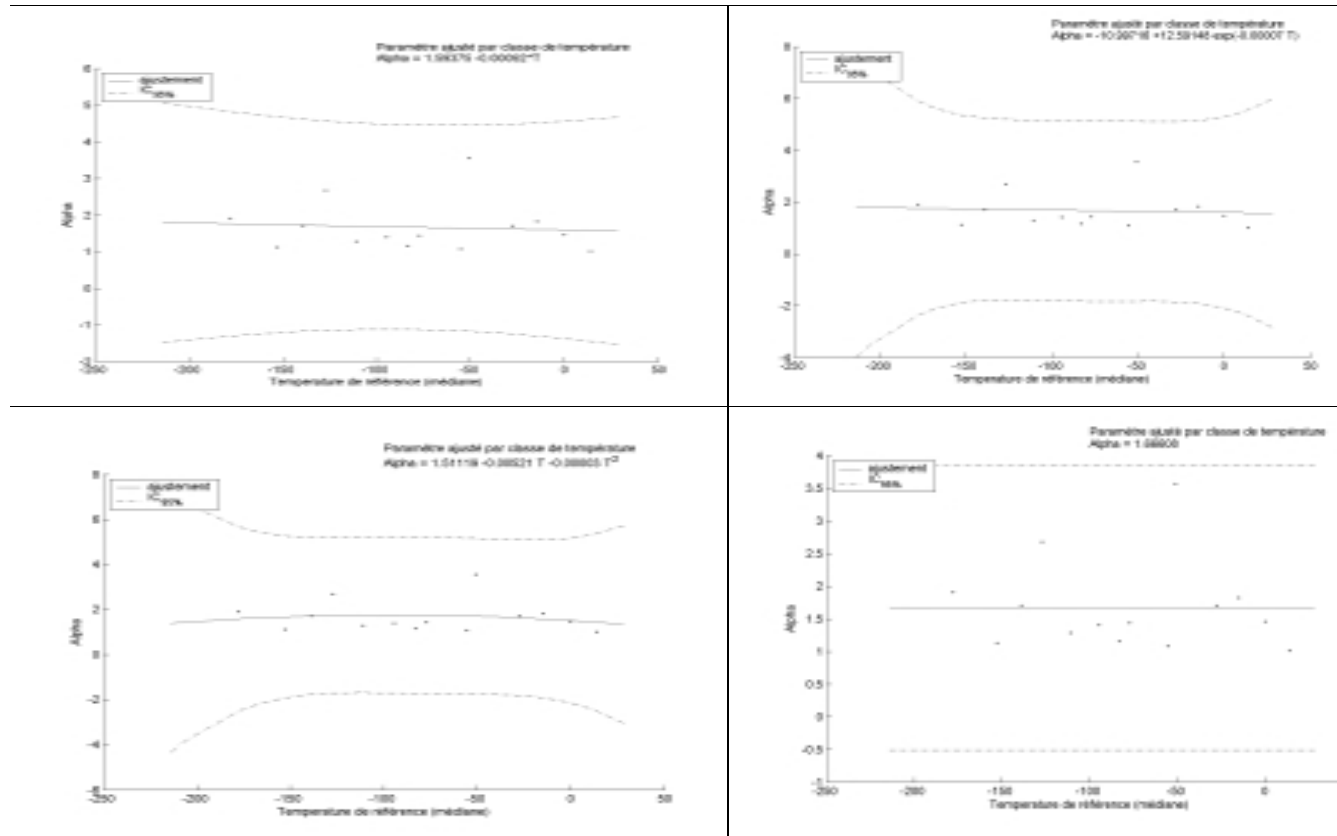
**Mutation** : tirage au hasard d'une nouvelle valeurs d'un gène pour un individu tiré lui même au hasard dans la population avec une Probabilité  $P_m$

**Optimisation locale** : recherche d'un optimum local dans la dimension d'un gène tiré au hasard

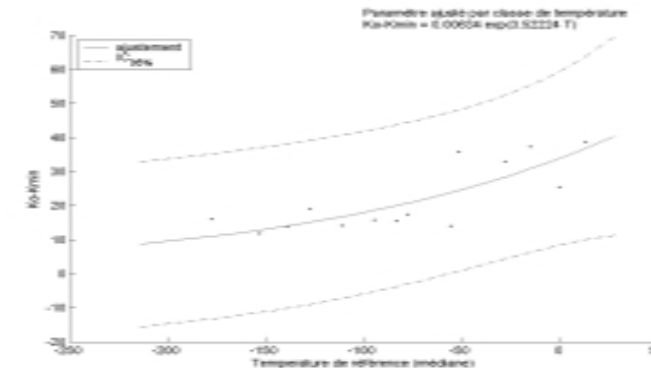
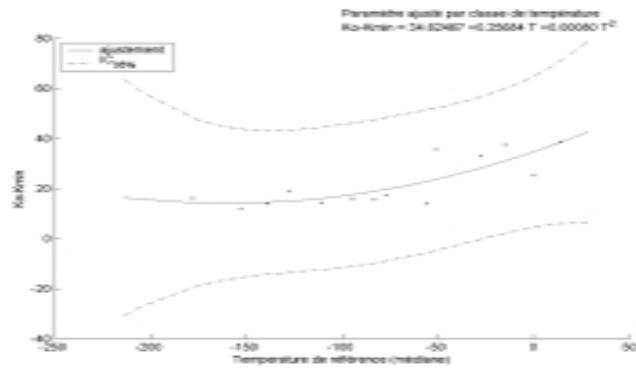
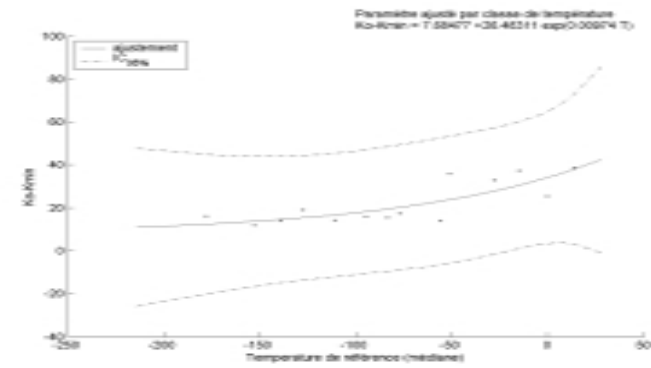
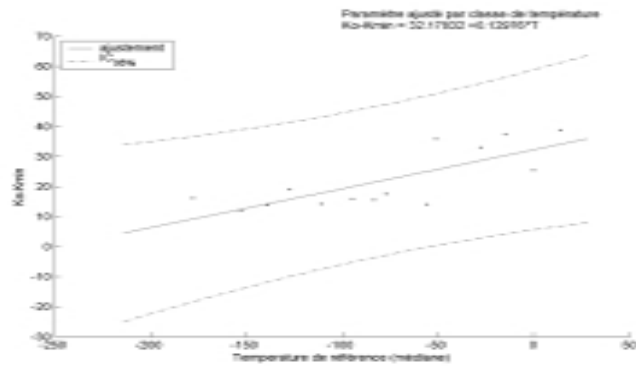
# Estimation de $K_{min}$



# Estimation de $\alpha$



# Estimation de $K_0 - K_{min}$

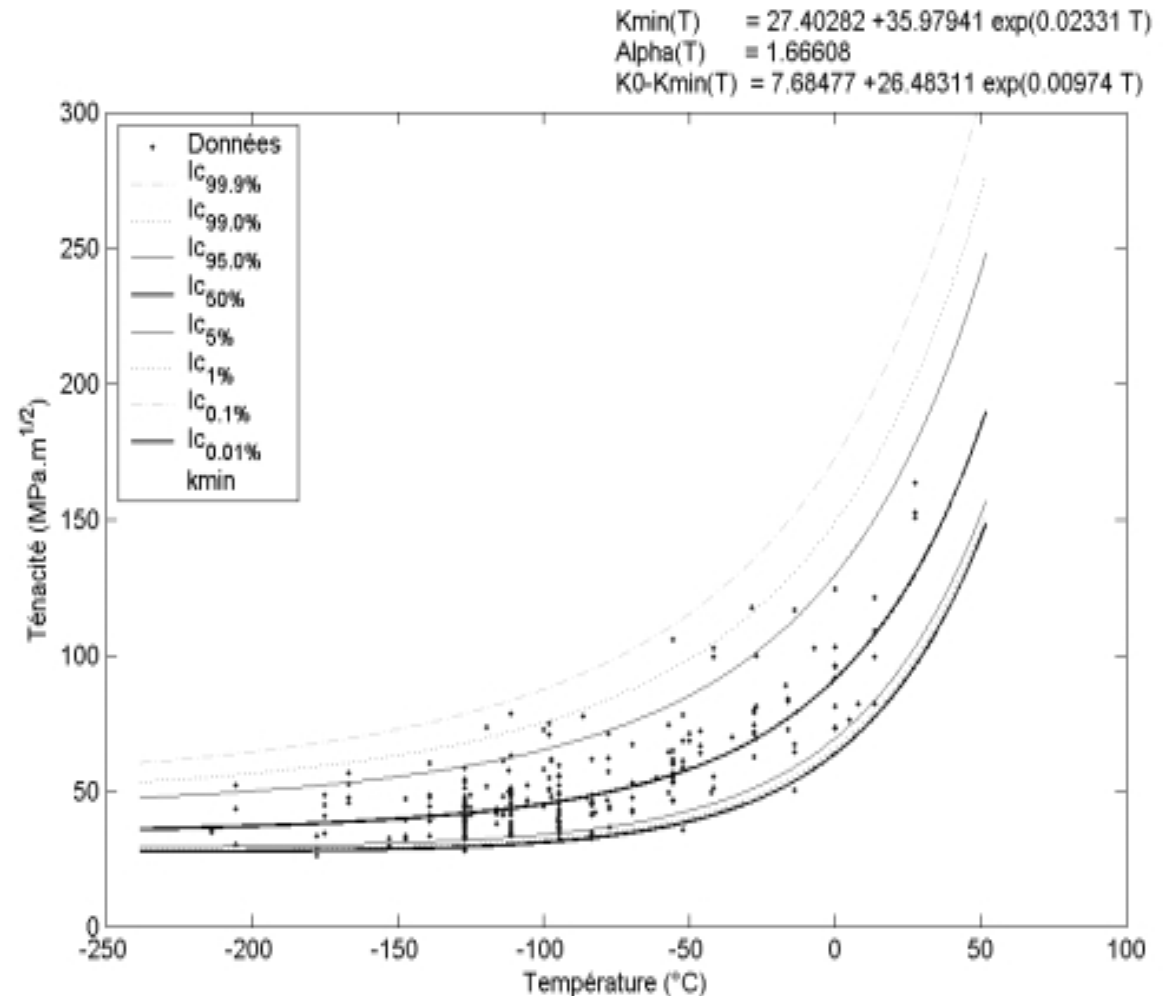




# Ajustement global



Cette figure présente différents fractiles de la distribution de Weibull obtenus avec les paramètres ajustés localement puis sur tout le domaine de température



# Ajustement global

---

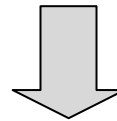


After estimating the parameters on each sample

→ global fitting of the parameters\* on the whole temperature domain.

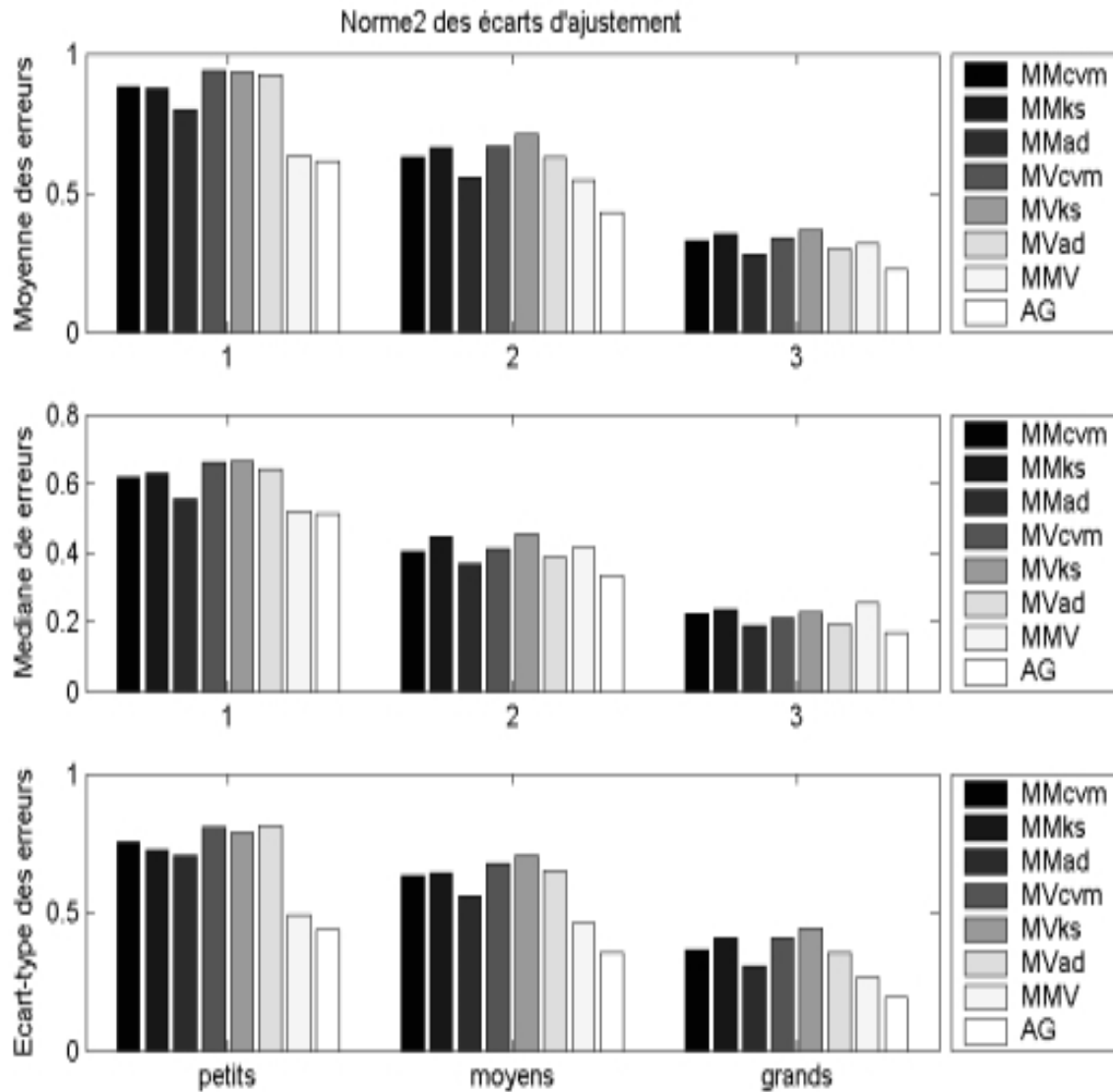
The two tests of goodness of fit ( $\chi^2$  and Cramer Von Mises) accept this preliminary fitting.

However this fitting is not fully acceptable because some data points are below the curve of  $K_{min}$ .



This is due to the procedure which cumulates the error of fitting of the different parameter.

# Quantitative comparison of the four methods



Comparison was made on simulated data and the error of fitting is defined by

$$\xi = \left( \frac{\mu_e - \hat{\mu}_e}{\mu_e} \right)^2 + \left( \frac{\alpha_e - \hat{\alpha}_e}{\alpha_e} \right)^2 + \left( \frac{\gamma_e - \hat{\gamma}_e}{\gamma_e} \right)^2$$

# Ajustement global direct sur tout le domaine de température

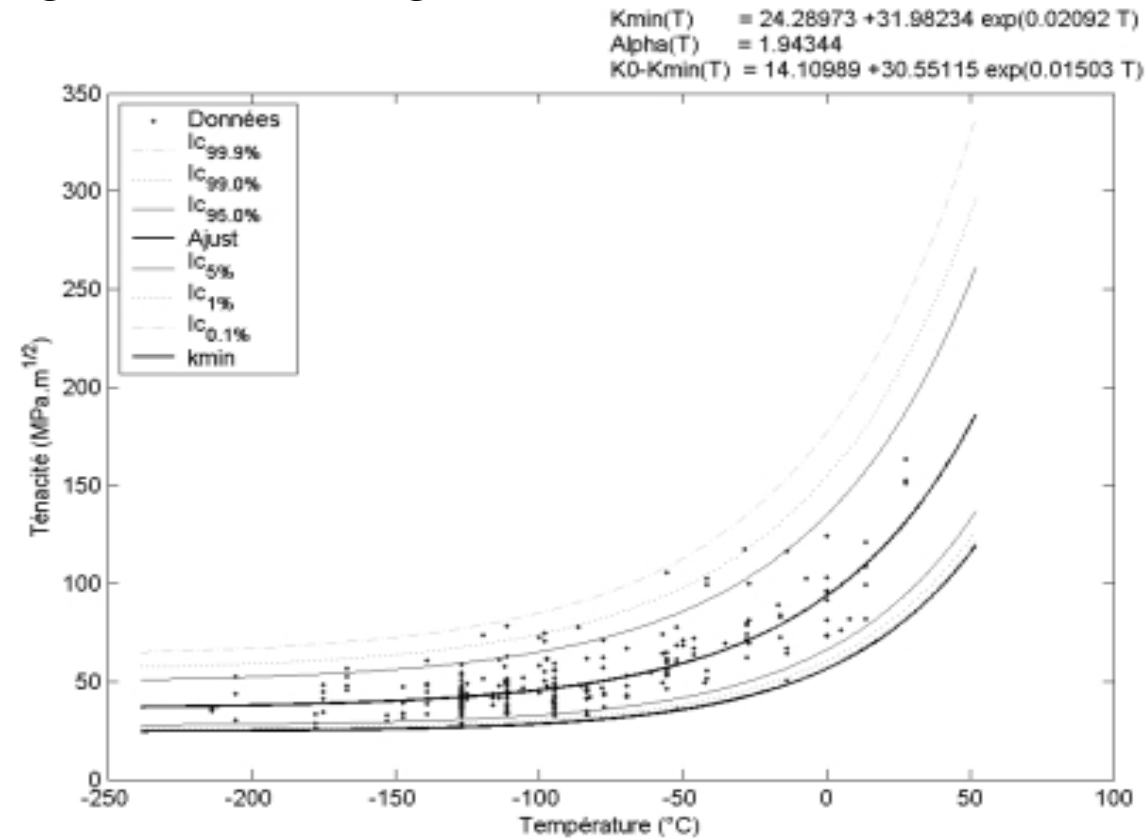


The direct global fitting is more accurate because there is no intermediary fitting of the functions and also because the genetic algorithm allows to add the supplementary constraint that any data points must be over the  $K_{\min}$  curve.

## Description

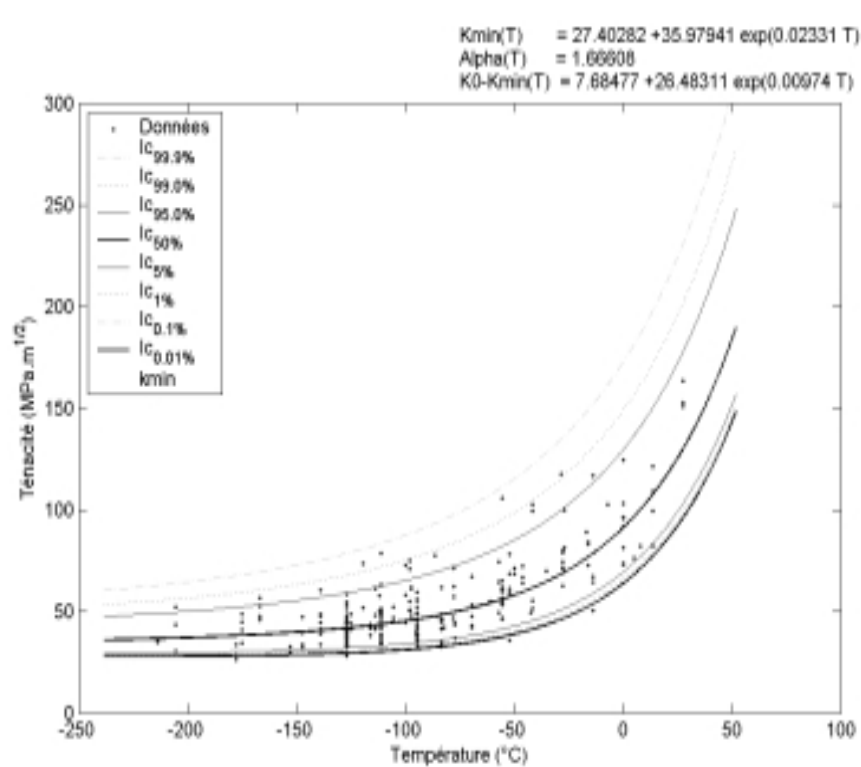
1. Preliminary fitting in order to choose the type of functions of temperature for the parameters of the Weibull distribution
2. Bootstrap calculation to estimate confidence intervals for each coefficients of the chosen function
3. Genetic algorithm calculation in order to find the optimal values of the coefficients inside the confidence intervals obtained in the step 2.

# Direct global fitting

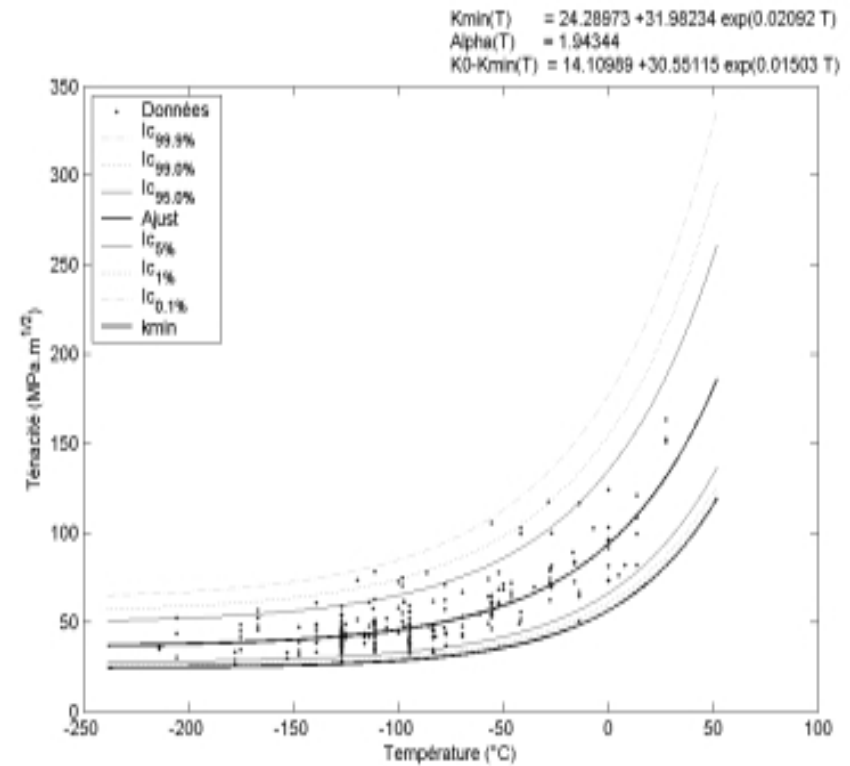


This figure presents different fractiles of the Weibull distribution obtained by this direct global fitting and the toughness data points. The two tests of goodness of fit ( $\chi^2$  and Cramer Von Mises) accept the global fitting. We can check that there is no data points below the  $K_{min}$  curve.

# Comparaison sous-échantillonnage/ajustement local et Ajustement global direct



Sous-échantillonnage/ajustement local



Ajustement global direct

## Conclusion

---



La méthodologie développée propose l'alternative :

1. Sous-échantillonnage/méthodes d'ajustement local
2. Méthode d'ajustement global direct plus précise et plus puissante.

Un logiciel a été développé pour implémenter ces méthodes dans l'environnement MATLAB : **Wolf 3**

# Perspectives

---



Étude de la robustesse de la méthodologie dans le cas d'un faible nombre de données

Quelle est la taille minimale de l'ensemble de données pour lequel la méthodologie reste robuste ?

Que faire pour des ensembles de données de taille inférieure à cette limite ?

- Bootstrap, ensemble flous.

Amélioration des techniques d'ajustement en queues de distribution

Validation des modèles dans les très bas fractiles de la distribution de ténacité.

- Test ET, GPD