Functional Weibull-based models of steel fracture toughness for structural risk analysis: estimation and selection

Nadia Pérot, Nicolas Bousquet

Functional Weibull-based models of steel fracture toughness for structural risk analysis: estimation and selection

Nadia Pérot[a], Nicolas Bousquet[b]

Summary: A key input component of numerous reliability studies of industrial components or structures, steel fracture toughness is usually considered as a random process because of its natural variability. Moreover, toughness presents a high sensitivity to temperature which also plays a fundamental role, as an environmental forcing, in such studies. Therefore a particular attention has to be paid to the assessment of its stochastic functional modelling, by means of a statistical analysis of indirect measures that suffer from heterogeneity and censoring. While a Weibull shape arising from statistical physics is recognized as the most relevant approach to represent local variability, the selection of best functional parameters (function of temperature) requires an accurate estimation and testing methodology. Its development is motivated by several limitations of the common statistical practices in the field of fracture toughness, which are related to data treatment and model selection. Illustrated by the exploration of a database feed by several European manufacturers or exploiters, this article establishes the main steps of such a methodology, implemented in a dedicated software tool.

Keywords: Weibull, Master Curve, Censoring, Genetic algorithm, Thickness correction, Model selection, fracture toughness, reliability

1. INTRODUCTION

Structural risk analysis (SRA) plays a key role in the management of passive and costly industrial components, especially those belonging to power production vessels or other pressurized systems for which safety must be guaranteed in critical situations. Such situations typically occur when a corrective action is performed that results in high-level stresses for the structure. For instance, Pressurized Water Reactor vessels must be cooled down by safety injections while still under pressure ; the injection of cold water causes a thermal shock transient which can weaken the integrity of the component [13]. SRA methodologies are mostly based on the simulation of degradation processes that have not been observed in reality or can not be reproduced in laboratory [25]. Roughly speaking, simulation models put *loads* L (including controllable actions) into competition with *capacities* R, and situations for which $L \geq R$ are defined as failures [14].

[a] *CEA Nuclear Energy Division, Centre de Cadarache, 13108, Saint-Paul-lès-Durance, France*
[b] *EDF Lab, 6 quai Watier, 78401 Chatou, France*

[*] *Correspondence to: nicolas.bousquet@edf.fr*

Amongst capacities the case of material fracture toughness is of major interest, since this property traduces the capability of the material to resist to pre-initiated crack propagation [19]. According to the weakest link (WL) physical theory [18], pre-initiated cracks correspond to heterogeneities within the crystal lattice of the material, and often arises from manufacturing defects. Such heterogeneities appear randomly, which is traduced by a natural variability of the fracture toughness (see Figure 1 for an illustration) and its modelling as a random variable in the dedicated literature [32]. Consequently, $L - R$ is randomized and a crucial reliability indicator is the probability $P(L > R)$ of a failure event. Numerous SRA methods deal with the computation of this probability, based on numerical exploration of the simulation model [34]. Steel being one of the most used materials in industry, the high sensitivity of its fracture toughness to temperature variations is representative of the statistical modelling difficulties encountered by reliability engineers [11]. Predicting how the toughness increases with temperature, from a brittle to a ductile nature, and summarizing this transition by a representative temperature [6], are two key issues of detailed SRA studies [38].

[Figure 1 about here.]

According to the WL theory, the brittle fracture toughness K_{IC} is explained at low temperature T by a functional Weibull model, popularized by [42] for a wide range of ferritic steels, and established as a US norm [2] under the so-called *Master Curve* (MC) denomination:

$$P(K_{IC}(T) < x) \quad = \quad 1 - \exp\left\{ - \left(\frac{x - K_{\min}}{K_0(T) - K_{\min}} \right)^{\alpha} \right\} \tag{1}$$

with shape parameter $\alpha = 4$, location parameter (or *brittle stage*) $K_{\min} = 20$ MPa.m$^{1/2}$ and with functional form

$$K_0(T) - K_{\min} \quad = \quad b_1 + b_2 \exp(-b_3 \cdot T). \tag{2}$$

The estimation of the parameter vector $\theta = (b_1, b_2, b_3) \in \mathbb{R}_*^+$ is conducted (usually using maximum likelihood techniques [46]) from fracture toughness observations produced from destructive tests on small-size specimen [7], that requires a so-called *thickness correction* to homogenize the corresponding observations. However, in the common assessment practices [23, 29, 37], several limitations of the MC model and methodological lacks interfere with an accurate use of statistical modelling, especially about the prediction of the brittle stage $K_{\min}$, which must be not overestimated as it resumes the minimal resistance to cold shocks.

1. The MC model and the value $\alpha = 4$ are relevant based on an assumption on the plasticity of cracks priming [43] and chemical homogeneity, which are not fully ensured in the case of welded components, and within all temperature ranges corresponding to experimental conditions. Consequently, the MC model threatens to be too rigid to explain the variability of observations, especially when the latter comes from various experiments conducted on specimen of close material grades by different laboratories. Despite several adaptations of methodologies of toughness quantification based on the

MC model (see for instance [47]), this lack of flexibility was still noticed by [22], from a modelling work guided by the EURO database originally used by [44], that aggregates steels from various European manufacturers. Alternatively, a recent competitor of the MC methodology, the Unified Curve method [27], offers a possible variation of the curve shape when the degree of embrittlement increases, but was criticized by [45] for its lack of universality. Adopting a universal encompassing approach, [22] provided a first answer to the issue of versatility by adding α to the vector θ of free parameters and proposing a statistical assessment based on maximum likelihood estimation (MLE). However, these authors missed the fact that adding this supplementary degree of freedom forbids to agglomerate heterogeneous observations directly using the specimen thickness correction proposed in the MC methodology, a key ingredient of the latter.

2. Experimental data, consisting in a transformation of priming measures on precracked test samples, present several degrees of validity fixed by norms [4, 20]. Especially, some of them correspond to limit (quasi-valid) cases when the level of energy used for the destructive test is too low or too high . Being currently rejected in the assessment methodologies of the MC model, these observations still yield relevant statistical censoring information, and should be used in the assessment of any toughness model.

Therefore the aim of this article is to provide a general methodology of statistical fracture toughness models based on functional Weibull forms, feed by heterogeneous databases, that: a) encompasses the MC methods using richer statistical models; b) solves the coherency problem raised by thickness correction when using such models; c) allow to incorporate the majority of experimental data and: d) provide adapted tools to model selection previous to the computation of functions of interest, as a brittle rupture reference temperature.

More precisely, the article is written as follows. Section 2 is dedicated to a brief presentation of the experimental context, a short reminder of the MC methodology and the proposed statistical methodology itself. It includes a data modelling step and the choice of appropriate functional forms $\{\alpha(T), K_{\min}(T), K_0(T)\}$ generalizing (1). The appropriate prior selection and statistical assessment of a class of toughness models, through the development of a specific genetic algorithm, is considered in Section 3. Section 4 describes the most adapted statistical methods for comparing the assessed models. The methodology is then applied to simulated and observed datasets, by means of the WOLF3 software [31]. Finally, main results, remaining issues and future research avenues are summarized and discussed in a dedicated section.

2. STATISTICAL MODELLING OF FRACTURE TOUGHNESS DATA

The realization of a fracture toughness random variable $K_{IC} = K_{i,T}$ at a given temperature T can be produced by destructive experiments based on a mechanical stress imposed on a pre-cracked specimen [47, 46]. The major feature of a specimen i is its thickness $B_{i,T}$, which typically evolves between 25 mm and 100 mm. Toughness is determined for a reference thickness B_0 (25 mm according to the ASTM norm [5]) and is assumed to follow, according to

the WL theory, the three-parameter Weibull distribution (1). The heterogeneity of specimen thicknesses can be discarded using the scale invariance property of the (reduced) two-parameter Weibull distribution, by the following transformation of the original samples:

$$K'_{i,T} = K_{\min} + (K_{i,T} - K_{min}) \times \left(\frac{B_{i,T}}{B_0} \right)^{\frac{1}{\alpha}}, \tag{3}$$

provided $(\alpha, K_{\min})$ are known. The first step of the MC methodology [35], assuming $\alpha = 4$ and $K_{\min} = 20$ MPa.m$^{1/2}$ is to produce this transformed sample. The estimation of $\{b_1, b_2, b_3\}$ follows. However, assuming an encompassing statistical framework for the MC model requires that $\{\alpha, K_{\min}\}$ should be estimated in parallel to $\{b_1, b_2, b_3\}$. While several authors (as [22]) prefer to make this correction before estimation, a more appropriate and fair estimation of $\theta = \{\alpha, K_{\min}, b_1, b_2, b_3\}$ should be based on maximizing the statistical likelihood defined, for one original data $k_{i,T}$, by

$$\begin{aligned} f_{K_{IC}}(k_{i,T}) &= \left(\frac{\alpha}{(K_0 - K_{min})(T)} \right) \times \left(\frac{k_{i,T} - K_{min}(T)}{(K_0 - K_{min})(T)} \times \left(\frac{B_{i,T}}{B_0} \right)^{\frac{1}{\alpha}} \right)^{(\alpha - 1)} \\ &\times \exp\left(- \left[\frac{k_{i,T} - K_{min}(T)}{(K_0 - K_{min})(T)} \times \left(\frac{B_{i,T}}{B_0} \right)^{\frac{1}{\alpha}} \right]^{\alpha} \right). \end{aligned} \tag{4}$$

The measurements generally considered as correct [20] are those referred to as K_{IC}, obtained by the procedure specified in ASTM E399-90, and the indirect K_{JC} elasto-plastic energy measurements which attempt to mitigate non-compliance with the linear constraints required by the mechanical theory (applied for obtaining K_{IC} to be valid).

In addition, empirical data can be obtained for different sample sizes and test temperatures which correspond to limit values (upper or lower bounds) for a *missing* toughness observation according to the classification by [20]. Such data can typically correspond to experiments conducted in the ductile range, without complete cracking, or, alternatively, by experiments "leading to large-scale yielding, exceeding the specimen's measuring capacity limit" [46]. Such data yield *censoring* information that is statistically relevant [3]. Most data referred to as K_{CM}, K_{CPM} and K_{MAX} in international nomenclatures [46, 51, 33] originate from quasi-valid experiments and may be considered as minimum limits (*right-censored*) for a missing toughness value. The likelihood contribution of a $k_{cm,i,T}$ (or $k_{jc-lim,i,T}$, etc.) value is then

$$P(K_{IC} > k_{cm,i,T}) = \exp\left(- \left[\frac{k_{cm,i,T} - K_{min}(T))}{(K_0 - K_{min})(T)} \times \left(\frac{B_{i,T}}{B_0} \right)^{\frac{1}{\alpha}} \right]^{\alpha} \right). \tag{5}$$

Conversely, other data referred as K_{JC-lim} correspond to going beyond the range of relevance

of toughness measurements and constitute upper limits (*left-censored*) for the expected toughness value. The statistical likelihood term of such a value obtained with a test specimen of thickness $B_{i,T}$ at a test temperature of T is then

$$P(K_{IC} < k_{jc-lim,i,T}) = 1 - \exp\left(- \left[\frac{k_{jc-lim,i,T} - K_{min}(T))}{(K_0 - K_{min})(T)} \times \left(\frac{B_{i,T}}{B_0} \right)^{\frac{1}{\alpha}} \right]^{\alpha} \right). \quad (6)$$

Incorporating both the thickness correction and the addition of censored values in the toughness statistical assessment is, by itself, an innovation in the field of fracture mechanics. Furthermore, more versatility can be given to the statistical model by considering several possible functional forms for the unknown parameter vector θ in function of T: apart from being constant, each component θ_k can be described as an increasing function of T (respecting the typical banana shape of toughness distribution, as in Figure 1). With $(a_k, b_k, c_k) \in \mathbb{R}_*^+$, the selected functions are:

1. linear: $\theta_k = a_k + b_k \times T$;
2. quadratic: $\theta_k = a_k + b_k T + c_k T^2$;
3. exponential: $\theta_k = a_k \exp(b_k T)$;
4. shifted exponential: $\theta_k = a_k + b_k \times \exp(c_k T)$.

Next section is dedicated to the prior selection of these functional models, for each dimension θ_k, then the overall fitting of the prior selected models using the likelihood maximum principle. Once the estimations conducted, a phase of posterior model selection is required to select the best candidate. This will be considered in Section 4.

3. PRIOR MODELS CLASS SELECTION AND STATISTICAL ASSESSMENT

The functional forms described above generate a wide range of possible encompassing statistical models that should be restrained before conducting parallel assessments. Only the most appropriate forms a priori must compete to explain the observations then be selected by statistical methods. Therefore a first step of the methodology, fully implemented within the WOLF3 software [31, 30], is to select these appropriate forms, using local estimation principles. Then an overall fitting is conducted. It should be noticed that, while the censored observations carry information to this second task, they are not used for the first one, since they cannot help to discriminate between forms unlike toughness data considered as correct.

3.1. Selection of appropriate functional models by local estimation

When toughness values are highly dispersed in relation to temperature, Weibull parameters must be estimated over a reduced temperature range and estimation must be conducted in relation to the reference temperature for this range. This reference temperature is either the mean temperature or the median temperature. In order to process the entire temperature

range of the database, two types of toughness data sub-sampling are carried out, based on a subdivision of the temperature range such that each sub-sample of data corresponds to the associated data in a sub-range of temperature: sequential sub-sampling and sliding sub-sampling. Weibull parameters are then estimated for each sub-sample. With the local estimates obtained, a functional temperature model can then be selected for each component of θ.

Sequential sup-sampling involves subdividing the temperature range I_T of the database into N separate consecutive sub-ranges$(I_i)_{i=1,...,n}$ with the same thermal amplitude ΔT. Alternatively, *sliding sub-sampling* involves subdividing the temperature range I_T of the database into N consecutive sub-ranges ΔT. We then have to build an initial temperature range I_1 starting with the lowest temperature T_1 of the temperature range up to temperature $T_1 + \Delta T$ and the next sub-range I_2 is obtained by sliding the sub-range I_1 by a shift of dT. Accordingly, $I_2 = [T_1 + dT, T_1 + dT + \Delta T]$. This operation is repeated until the sub-range I_n reaches the maximum temperature of I_T.

Local estimations follow. For a fixed T, Weibull parameters $\alpha, K_{\min}$ and $K_0 - K_{\min}$ are estimated for each of the N_D sub-ranges obtained by sequential or sliding sub-sampling. The following methods are considered, that take into account the thickness for each toughness value.

1. *The moment method*: for all α (shape parameter) varying in the interval $[a; b]$ with a step h, the estimation of $K_0 - K_{min}$ and K_{min} is conducted by the method of moments [12]. Only the triplets $(\alpha, (K_0 - K_{min})^*, K^*_{min})$ having a physical sense are retained;
2. *The maximum likelihood method* consists in searching the values of the parameters that maximize the likelihood function, for the three-parameter Weibull distribution: following Smith and Lawless' advice [39, 24] for all K_{min} (position parameter) varying in the interval $[a; b]$ with a step h, the estimation of the parameters α and $K_0 - K_{min}$ is conducted by the maximum likelihood method. The triplet $(\tilde{\alpha}, (K_0 - K_{min})^*, K^*_{min})$ which maximizes the complete likelihood is selected;
3. *A hybrid method* for the Weibull distribution which combines the previous ones: the K_{min} parameter is estimated by the moment method and then others are estimated by the maximum likelihood method.

The moment and maximum likelihood methods produce several triplet solutions which are hierarchized using Cramer-Von Mises, Kolmogorov-Smirnov and Anderson-Darling criteria [40]. In each test, the triplet minimizing the associated statistical value is selected. Finally, N_D local estimates $(\alpha^*, (K_0 - K_{min})^*, K^*_{min})_{i=1,...,N_D}$ are obtained, where N_D is the number of sub-samples. These samples allows for a graphical evaluation of the relevance of functional forms described in previous section, as well as fitting using usual least-square techniques.

3.2. Maximum likelihood overall estimation

This second step consists in calculating the best functional estimates $(\alpha^*(T), (K_0 - K_{min})^*(T), K^*_{min}(T))$ by maximizing the overall likelihood defined as

$$\begin{aligned}\mathcal{L}_{K_{min},\alpha,(K_0-K_{min})}(\mathbf{k}) &= \mathcal{L}_1(K_{min},\alpha,(K_0-K_{min}))(\mathbf{k_{1C}}) \\ &\quad \times \mathcal{L}_2(K_{min},\alpha,(K_0-K_{min}))(\mathbf{k_r}) \\ &\quad \times \mathcal{L}_3(K_{min},\alpha,(K_0-K_{min}))(\mathbf{k_l}),\end{aligned}$$

where $\mathbf{k} = \{\mathbf{k_{1C}}, \mathbf{k_r}, \mathbf{k_l}\}$ denotes all data produced by experiments, $\mathbf{k_{1C}}$ being the regular toughness data while $\mathbf{k_r}$ (*resp.* $\mathbf{k_l}$) are the right-censored (*resp.* left-censored) observations. Assuming the independence of measurements, the corresponding likelihoods $\mathcal{L}_1$, $\mathcal{L}_2$ and $\mathcal{L}_3$ are products of terms described in Section 2. Replacing α, $(K_0 - K_{min})$ and K_{min} in $\mathcal{L}_{K_{min},\alpha,(K_0-K_{min})}(\mathbf{k})$ by functionals $\alpha(T)$, $(K_0 - K_{min})(T)$ and $K_{min}(T)$ parametrized by $(a_i, b_j, c_l)_{i\in I, j\in J, l\in L}$, the optimisation problem becomes to estimate

$$\begin{aligned}(a_i, b_j, c_l)^*_{i,j,l} &= \arg\max \log \mathcal{L}_{K_{min}(T),\alpha(T),(K_0-K_{min})(T)}(\mathbf{k}) \qquad (7)\\ &\text{under the constraints} \\ \alpha(T) &> 2 \quad \forall T, \\ (K_0 - K_{min})(T) &> 0 \quad \forall T, \\ 0 < K_{min}(T) &< \min_{(i,T)\in\mathbf{k_{1C}}} k_{(i,T)}.\end{aligned}$$

Genetic algorithms [28] are general-purpose search algorithm based upon the principles of evolution in nature (permanent adaptation). They can be applied to a wide variety of optimisation problems [16] and appeared of good relevance to solve (7). For nonregular models, compact sets for the variation ranges of coefficients to be estimated usually appear necessary to obtain non-degenerate and consistent MLE [39], in addition of other constraints (for instance, the true value of α should be upper than 2 when considering a nonfunctional three-parameter Weibull distribution [50]). More generally, it is a prerequisite for starting genetic algorithms. In the WOLF3 software, such ranges can be directly informed by the user or provided by a bootstrap algorithm, described hereafter. In numerical experiments these ranges were found to contain the true value for each coefficient, but obviously this cannot be guaranteed in all situations.

Non-parametric bootstrap calibration of variation ranges.

1. Sample with replacement N_{Boot} datasets $\{d_1, \ldots, d_{N_{Boot}}\}$, each of size N, amongst the N (uncensored) original toughness observations;
2. For replicate d_i, produce $k_i < N$ local estimations of $\{\alpha, K_{\min}, K_0 - K_{\min}\}$ and fit the parametric model chosen for each parameter;
3. Estimate empirically the quartiles $\{q_{k,1}, q_{k,2}, q_{k,3}\}$ from the $N_{Boot}-$sized sample of estimates for each coefficient θ_k;

4. Calibrate the range for θ_k as

$$[q_{k,2} - 3 \cdot (q_{k,3} - q_{k,1}) \quad q_{k,2} + 3 \cdot (q_{k,3} - q_{k,1})] \,.$$

The genetic algorithm proposed in this article is based on the definition of a population of N_p individuals. Each individual represents a point in the space of states which means a candidate solution. It is characterized by a set of genes (the values of the variables to be estimated) and a fitness function (the value of the criterion to be optimized). The algorithm then generates populations at each iteration, on which selection and mutation processes are applied, the purpose of which is to ensure that the space of states is efficiently explored. The evolution of all individuals over several generations leads to the optimum states for the relevant optimisation problem.

The entire process is carried out for a constant population size and each iteration is referred to as a generation by analogy with genetics. A population initially built by random sampling evolves from a k generation to a $k+1$ generation by applying the following operations to the individuals:

- **Evaluation:** in calculating the fitness of the individuals to the problem solution, $\mathcal{L}(K_{min}, \alpha, (K_0 - K_{min}))(\mathbf{k})$ is calculated for genes corresponding to the coefficient values of functional $\alpha(T)$, $(K_0 - K_{min})(T)$ and $K_{min}(T)$. However, if the constraints of Problem (7) cannot be satisfied due to the genes of an individual, its fitness is fixed at $-\infty$.
- **Selection:** designates the individuals best adapted to survive and transmit their genes in relation to their fitness.
- **Crossing:** allows genes from two individuals to be mixed to give two offspring individuals intended to replace them.
- **Mutation:** modifies a gene for certain randomly sampled individuals.

The algorithm can be stopped when the population ceases to evolve or for a fixed number of generations. The individual showing the greatest fitness in the final population then corresponds to a solution to the problem.

4. FINAL MODEL SELECTION

It must be noticed that the class of models defined in Section 2 encompasses the MC model and other nested models, which implies that statistical testing between assessed models can be conducted using powerful tools as likelihood ratio tests [17]. See Appendix B for an example considering several relaxations of the MC model. Structural differences between linear and exponential functions require that other tools of model selection be used, as Aikake (AIC) and Bayesian (BIC) Information Criteria [1, 36] adapted to censored situations [41, 26**?**]. In practice, AIC should be preferred since it was conceived to be efficient in a finite list of approximate models, optimizing the trade-off between bias and

variance [8], while BIC (known to select models with smaller dimension than AIC, possibly underfitted) is consistent in the sense it selects (asymptotically) the true model in a class if it is assumed to be unique and belong to this class [9]. In addition, a conditional χ^2 test was specifically developed to help selecting the most relevant model among all tested ones, from the uncensored observations. A summary of this procedure is presented below, while the details can be found in [30].

Considers a set $\{K_{IC}^{(q)}, T^{(q)}\}_{q=1,\cdots,n}$ of toughness measures and indexation temperatures and assume that all data pairs are mutually independent. The fitness of a traditional χ^2 test [48] to this set of pair of variables requires a subdivision of the space of the variable $K_{IC}|T = x$ into L classes S_l ($cf.$ Figures 2 and 3). Then the $L-$sized observation vector N_{obs} is compared to the $L-$sized theoretical vector N_{exp}, both defined as:

$$N_{obs} = \begin{pmatrix} \sum_{q=1}^{n} \mathbb{1}_{K_{IC}^{(q)} \in S_1} \\ \vdots \\ \sum_{q=1}^{n} \mathbb{1}_{K_{IC}^{(q)} \in S_L} \end{pmatrix} \quad \text{and} \quad N_{exp} = \begin{pmatrix} \sum_{k=1}^{K} n_k q_{1,k} \\ \vdots \\ \sum_{k=1}^{K} n_k q_{L,k} \end{pmatrix}$$

where, $\forall l \in \{1, ..., L\}$ and $\forall k \in \{1, ..., K\}$,

$$q_{l,k} \;=\; \mathbb{P}\left(K_{IC} \in S_l | T = x_k\right),$$

and n_k is the number of times when $T = x_k$.

[Figure 2 about here.]

[Figure 3 about here.]

Finally,

$$Z \;=\; N_{obs} - N_{exp} \;\sim\; \mathcal{N}_L(0, \Gamma)$$

with

$$\Gamma = n \begin{pmatrix} \sum_{k=1}^{K} n_k q_{1,k}[1-q_{1,k}] & -\sum_{k=1}^{K} n_k q_{1,k} q_{2,k} & \cdots & -\sum_{k=1}^{K} n_k q_{1,k} q_{L,k} \\ -\sum_{k=1}^{K} n_k q_{2,k} q_{1,k} & \sum_{k=1}^{K} n_k q_{2,k}[1-q_{2,k}] & \cdots & -\sum_{k=1}^{K} n_k q_{2,k} q_{L,k} \\ \vdots & \vdots & \ddots & \vdots \\ -\sum_{k=1}^{K} n_k q_{L,k} q_{1,k} & -\sum_{k=1}^{K} n_k q_{L,k} q_{2,k} & \cdots & \sum_{k=1}^{K} n_k q_{L,k}[1-q_{L,k}] \end{pmatrix}$$

where K is the number of distinct values of T.

$$U\Sigma^{-1}U' \sim \chi^2_Q$$

Noting Z^* the vector containing only the first $L-1$ components of Z and Γ^* its covariance matrix (i.e. the matrix Γ without the last column and the last line), it comes, under the null hypothesis H_0 that the tested model is true:

$$Z^*(\Gamma^*)^{-1}Z^{*'} \sim \chi^2_{L-1}$$

and H_0 will be rejected at threshold ϵ if the test statistic $Z^*(\Gamma^*)^{-1}Z^{*'}$ exceeds the percentile $\chi^2_{L-1}(1-\epsilon)$.

5. NUMERICAL EXPERIMENTS

The following numerical experiments are conducted from a so-called EURO database of 849 real steel (16 MDN5) toughness measurements (Figure 1), aggregating data from different European manufacturers or exploiters (SIEMENS, EDF, CEA, FRAMATOME, AEA). Various versions of this database, according to whether or not it includes toughness data considered as quasi-valid, non-valid but informative or poorly informative, were used in [20] and [22].

Sequential sub-sampling based on a $20°C$ width, involving a minimum of 20 data per sample, was conducted on the regular data (Figure 4). By local estimation using the method of moments on each sub-sample, $N = 7$ triplets $(K^{(i)}_{min}, \alpha^{(i)}, K^{(i)}_0 - K^{(i)}_{min})_{i=1,\ldots,N}$ are assessed. In Figure 4, the local estimates are fitted by a linear function for $K_{min}(T)$, a constant value for $\alpha(T)$ and a shifted exponential function for $(K_0 - K_{min})(T)$.

[Figure 4 about here.]

An overall ML estimation was conducted, gradually increasing the data size. The results are set out in Table 1. Divergences standardised between empirical and theoretical quantiles are also traced (QQ-plots, cf. Figures 7 and 8) and summarised in the same table, focusing separately on the high and low sections of the transition curve. Taking into account censoring information, the estimated model provides mean deviation between empirical and theoretical quantiles which is almost 6% better for all quantiles together, almost 9% better for high quantiles (75%-99%) add almost 2% better for low quantiles (0.1%-25%) in relation to the dispersion found when censoring is not taken into account. Hence, including the censoring information increases the information on the toughness model parameters coherently with the structure of the model. The relevance of the estimations summarized in Table 1 is verified by performing simulated tests in next two subsections.

[Table 1 about here.]

[Figure 5 about here.]

5.1. Initial experiments

30 sets of 849 values each were simulated from the following estimates, which are very similar to those obtained from the EURO database: $K_{\min} = 20$, $\alpha = 3$ and $K_0(T) - K_{\min} = 0.004 + 424 \cdot \exp(0.01472 \cdot T)$. The test temperatures are the same as those in the original dataset and every dataset complies with the proportion of censored data in the latter (4.4% right-censored, 59% left-censored). Additional details about the features of the simulation process are given in Appendix A.

Table 2 summarizes the estimation results. All simulated values of the parameters are located within the standard confidence ranges built from the estimators. Note in particular that the standard deviation of the estimator on the ordinate at the origin of $K_0(T) - K\min(T)$ puts into perspective the difference observed between the simulated value and the average estimate obtained from the 30 samples. The estimation procedure presented in this article and implemented in the WOLF3 software [31, 30] thus provides relevant results and, in particular, gives a good estimate of the brittle phase (the ordinate at the origin of $K_{\min}(T)$).

[Table 2 about here.]

5.2. Subsequent experiments

Secondly, testing is required to establish whether a more complex model encompassing the traditional MC is accurately estimated if the simulated data actually come from a MC: the additional parameters must be estimated at 0 or near to 0 and the more flexible models must adopt a similar behavior. Accordingly, by Ockham's rule of least complexity and on the basis of statistics from traditional testing procedures (e.g., AIC), the simplest model most used in practice should be selected.

The simulation parameters are therefore chosen as follows: $K_{\min} = 20$, $\alpha = 4$ and $K_0(T) - K_{\min} = 50 + 200 \cdot \exp(0.002 \cdot T)$, and the test temperatures and censoring values are selected as previously. Again, 30 independent datasets are simulated, of which an example is shown in Figure 10 (in Appendix). The estimation results are shown in Table 3. The assumption $\alpha = 4$, essentially characteristic of the Master Curve, is applied by these models. The low level of linearity noted for $K_{\min}$ has little effect on the shape of the curve and, for its part, the parameter that determines the exponential shape is well-estimated. Additional complexity (passing from a linear model to a shifted exponential model for $K_{\min}$ is logically manifested in increased estimated standard deviation values. However, through limited development, the low value of the exponential coefficient of $K_{\min}$ allows an equivalent linear model to be obtained and the brittle phase to be quantified between 20 and 25 $MPa.\sqrt{m}$. As expected, this brings us back to the main features of the Master Curve.

[Table 3 about here.]

5.3. Testing the Master Curve in the EURO database

Finally, the statistical relevance of the classic MC model over the motivating EURO dataset is compared to the other possible models defined by the encompassing framework. Results of fitting are summarized on Tables 4 (including the MC model) and 5. The AIC criterion is defined classically [1] as the penalization of twice the maximized negative loglikelihood:

$$\text{AIC} = -2 \log \mathcal{L}_{\hat{\theta}}(\mathbf{k} + 2d$$

in the uncensored case, with d the dimension of the model, $\mathcal{L}$ its likelihood and $\hat{\theta}$ the MLE of the unknown parameter vector θ, and

$$\text{AIC} = -2 \log \mathcal{L}_{\hat{\theta}}(\mathbf{k} + d + \text{tr}\left(I_{\text{all},\hat{\theta}} I^{-1}_{\text{incomp},\hat{\theta}}\right) \tag{8}$$

in the censored case, following [**?**], where $(I_{\text{all},\theta}, I_{\text{incomp},\theta})$ are the Fisher information matrices for the complete data and incomplete data, respectively defined by (for a single observation k)

$$I_{x,\theta} = -\int \mathcal{L}^x_\theta(k) \frac{\partial^2 \log \mathcal{L}^x_\theta(k)}{\partial\theta\partial\theta^T}\, dk,$$

where $x \in \{_{\text{incomp}}, _{\text{all}}\}$ and where $\mathcal{L}^{\text{incomp}}_\theta$ is either the density, the survival of the cumulative distribution function in k, while $\mathcal{L}^{\text{all}}_\theta$ is only the density of k. For these functional models these information quantities can be empirically computed from the observations. It was most often observed that, for the considered dataset, the extreme-right penalty in (8) was very close to d. This formulation means that a model with a low AIC value is considered to explain better the observations than a model with a high AIC value.

The results first confirm the necessity of relaxing the rigidity of the MC model, by considering more parameters are unknown a priori and conducting statistical estimation.

A negligible $p-$value of the χ^2 test (namely, a negligible probability of observing the test statistic under the assumption of the model in consideration), associated to a high AIC value, highlight that the MC model is comparable, in terms of explicative power, to basic models based on linear functionals only. The sensible gap in terms of AIC due to the insertion of a shifted exponential model with unknown parameters, for $K_{\min}$ or $K_0 - K\min$ (provided other parameters are unknown too), indicates that a good strategy for selecting a relevant model should at least take account of this criterion. Using a LRT test described in appendix B, it is possible to refine the diagnostic about the MC model: considering the relaxed MC model with unknown (but constant) $K_{\min}$ and α (with operational constraints $K_{\min} \geq 10$ and $\alpha > 2$), the observed statistic is $R_{4,1} \simeq 0.04923$ which is of the same order than the 5%-order percentile of the mixture of Dirac and χ^2 distributions ($\simeq 0.05375$).

If the quadratic evolution of $K_0 - K_{\min}$ seems to be the most relevant from the AIC viewpoint at the light of the results provided on both tables, a quick look on the corresponding figures (Figures S-15 and S-21 in Supplementary Online Material (SOM)) is enough to discard such a model from a physical point of view (no obversation is plausible at low temperature close to 100 MPa.$\sqrt{m}$). Rather, a good trade-off between statistical fitting and physical plausibility is provided by Models (14) and (17). The estimation of parameters (see Figures S-27, S-28 and S-33 in SOM) shows that the quadratic and exponential coefficients of $K_{\min}$ take most often very small values, and that these models can be easily derived (by Taylor expansion around 0) in the simpler form of Model (12), which is our final choice for this dataset. Note that the standard MC model and Model (12) strongly differ by their derived value of the reference temperature (gap $\sim 10°C$). Another important consequence is that the brittle stage $K_{\min}$ is increasing with the temperature. Such a result appears to be useful for risk engineers who would be able to define sensitivity analyses and margin assessments with respect to the conservative MC model.

Finally, it must be noticed that accounting for censored values can have a more sensible effect on the estimation of unknown parameters and (as expected) on the reference temperature, traduced by a possible difference of several degrees, than on the model selection itself.

[Table 4 about here.]

[Table 5 about here.]

6. DISCUSSION

This article presents a statistical methodology of estimation and selection of a class of steel fracture toughness models encompassing the celebrated Master Curve. Its implementation within a dedicated software was thought to simplify its use by reliability engineers. The common practice of this engineering field, as crude homogenization of experimental data and putting aside nonregular observations, as well as the practical necessity of using more flexible models than the Master Curve, motivated this work. An immediate benefit of

improving the statistical modelling of steel fracture toughness is improving the knowledge of the brittle stage and the brittle-ductile transition temperature range. While the brittle stage appears as a penalizing factor in structural reliability studies, the reference temperature can be used to hierarchize steels and compare steel structures.

Another interest of this refined modelling is guiding the design of new destructive experiments (while the use of censored observations yields supplementary information that, conversely, should diminish the necessity of such experiments). Indeed, designing these experiments is realizing a trade-off between costs and statistical information gain, through the use of cost functions and information measures integrated over the expected distribution of toughness [15]. In a Bayesian perspective, a prior model recognized as "the best on the market" can be used to derive accurate distributions for the coefficients and delimit the most informative ranges of temperature to explore, under fixed budget, to improve significantly the robustness of the statistical modelling [10]. This will be the subject of a future work.

This methodology remains clearly opened to improvements. First, the selection of functionals based on local estimations may suffer from a lack of estimated parameter values if the temperature ranges are chosen too wide. Using nonregular ranges to gain estimations may distort the estimated shapes and bias the selection of these functionals. Nonparametric tests could besides be adapted to provide objective help to this selection. Second, the overall optimisation task could probably be improved by taking account of the missing data structure due to the presence of censoring, using multiple imputation methods or data augmentation methods. Finally, the use of sensitivity analysis techniques [21] could be helpful for comparing the robustness of several assessed models, in complement to classic criteria, and improving the confidence that may be placed in the modelling of this very influential input of structural reliability studies.

7. ACKNOWLEDGMENTS

The authors gratefully thank Patrick Todeschini, Émilie Dautrême (EDF Lab) and Michel Marquès (CEA) for their help and advices during the work that guided the redaction of this article.

REFERENCES

[1] H. Aikake. A new look at the statistical model identification. IEEE Transactions on Automatic Control, 19(6):716–723, 1974.

[2] Anonymous. Guidelines for application of the master curve approach to reactor pressure vessel integrity in nuclear power plants. Technical report, AIEA, Technical report series Number 429, 2003.

[3] Anonymous. NIST/SEMATECH e-Handbook of Statistical Methods. National Institute of Standards and Technology: *http://www.itl.nist.gov/div898/handbook/*, 2012.

[4] ASTM. E399-90: Standard Test Method for Plane-Strain Fracture Toughness of Metallic Materials. Annual Book of ASTM Standards. American Society for Testing and Materials International, 1997.
[5] ASTM. E1921-12: Standard Test Method for Determination of Reference Temperature, To, for Ferritic Steels in the Transition Range. Annual Book of ASTM Standards. American Society for Testing and Materials International, 2011.
[6] ASTM. Test method for determination of reference temperature, t0, for ferritic steels in the transition range. Technical report, ASTM International, 2015.
[7] AWS. Best practices: destructive testing for material toughness. Inspection Trends, American Welding Society, pages 30–31, 2008.
[8] L. Birgé and P. Massart. Gaussian model selection. Journal of the European Mathematical Society, 3:203–268, 2001.
[9] K.P. Burnham and D. Anderson. Model selection and multi-model inference. New York: Springer-Verlag, 2002.
[10] K. Chaloner and I. Verdinelli. Bayesian experimental design: A review. Statistical Science, 10:273–304, 1995.
[11] T. Chandra, K. Tsuzaki, M Militzer, and C. Ravindran. Microstructure - texture related toughness anisotropy of api-x80 pipeline steel. Advanced Materials Research, 15-17:840–845, 2006.
[12] D. Cousineau. Fitting the three-parameter weibull distribution: Review and evaluation of existing and new methods. IEEE Transactions on Dielectrics and Electrical Insulation, 16:281–288, 2009.
[13] M. Deprost. Nuclear reactors: monitoring steel behaviour. Enviscope, November, 2010.
[14] A.C. Estes and D.M. Frangopol. Load rating versus reliability analysis. Journal of Structural Engineering, 131(5):843–847, 2005.
[15] J. Gladitz and J. Pilz. Construction of optimal designs in random coefficient regression models. Statistics, 13:371–385, 1982.
[16] D. E. Goldberg. Gentetic algorithm in search, optimzation, and machine learning. Addison Wesley, 1989.
[17] C. Gourerioux and A. Monfort. Statistique et modèles économétriques. Economica, Paris, 1996.
[18] A.M. Hasofer. A statistical theory of the brittle fracture of steel. International Journal of Fracture, 4:439–452, 1968.
[19] P. Haušild, I. Nedbal, C. Berdin, and C. Prioul. The influence of ductile tearing on fracture energy in the ductile-to-brittle transition temperature range. Materials Science and Engineering, 335:164–174, 2002.
[20] B. Houssin, R. Langer, D. Lidbury, T. Planman, and K. Wallin. Unified reference fracture toughness design curves for rpv steels. Technical report, EE/S.01.0163 Rev. B Final Report, 2001.
[21] B. Iooss and P. Lemaître. A review on global sensitivity analysis methods. In: Uncertainty Management in Simulation-Optimization of Complex Systems, G. Dellino, C. Meloni (eds), pages 101–122, 2015.
[22] F. Josse, Y. Lefebvre, P. Todeschini, S. Turato, and E. Meister. Statistical analyses for probabilistic assessments of the reactor pressure vessel structural integrity: Building a master curve on an extract of the "euro" fracture toughness dataset, controlling statistical uncertainty for both mono-temperature and multi-temperature tests. Proceedings of the 14th International Conference on Nuclear Engineering, Miami, Florida, 2006.
[23] S.S. Kang, S.H. Chi, and J.H. Hong. Statistical evaluation of fracture characteristics of rpv steels int the ductile-brittle transition temperature region. Journal of the Korean Nuclear Society, 30:364–376, 1998.
[24] J.F. Lawless. Statistical Models and Methods for Lifetime Data (3rd edition). New York: John Wiley and Sons, 2003.
[25] M. Lemaire, M. Pendola, and J.-C. Mitteau. Structural reliability. Wiley-ISTE, 2009.
[26] H. Liang and G. Zou. Improved aic selection strategy for survival analysis. Computational Statistics and Data Analysis, 52:2538–2548, 2008.
[27] B.Z. Margolin, A.G. Gulenko, V.A. Nikolaev, and L.N. Ryadkov. A new engineering method for

prediction of fracture toughness temperature dependence for pressure-vessel steels. Strength of Materials, 35:440–457, 2003.
[28] M. Mitchell. An Introduction to Genetic Algorithm. MIT Press, 1996.
[29] R. Moskovic. Modelling of fracture toughness data in the ductile to brittle transition temperature region by statistical analysis. Engineering Fracture Mechanics, pages 511–530, 2002.
[30] Bousquet N. Pérot, N. and M. Marques. Method for determining the strength distribution and the ductile-brittle transition temperature of a steel product subjected to thermal variations. European Patent Register. . Proposition 15718909.3, April 29th, 2015, 2015.
[31] N. Pérot and M. Marques. Wolf3: A statistical software for refined modelling of fracture toughness data indexed by temperature. In: Safety, Reliability, Risk and Life-Cycle Performance of Structures and Infrastructures, G. Deodatis , B.R. Ellingwood and D.M. Frangopol (eds), Proceedings of ICOSSAR 2013, New York, USA, CRC Press., 2014.
[32] A. Pineau. Development of the local approach to fracture over the past 25 years: theory and applications. Anales de la Mecánica de Fractura, 1:9–23, 2007.
[33] Y. Quemener, H. Chien-Hua, and L. Chi-Fang. Assessment of critical fatigue crack length considering the fracture failure of ship longitudinal members. Proceedings of the 27th Asian-Pacific Technical Exchange and Advisory Meeting on Marine Structures (TEAM), Sep. 9-12, Keelung, Taiwan, 2013.
[34] G. Rubino and B. Tuffin. Rare Event Simulation using Monte Carlo Methods. John Wiley, 2009.
[35] I. Sattari-Far and K. Wallin. Application of master curve fracture toughness methodology for structural integrity of nuclear components. SKI Report 2005:55, October 2005, 2005.
[36] Gideon E. Schwarz, H. Estimating the dimension of a model. Annals of Statistics, 6:461–464, 1978.
[37] S.Y. Shin, B. Hwang, S. Kim, and S. Lee. Fracture toughness analysis in transition temperature region of apix70 pipeline steels. Material Science and Engineering, 429:196–204, 2006.
[38] H. Sieurin. Fracture toughness properties of duplex stainless steels. Technical report, Ph.D. thesis, Royal Institute of Technology, Department of Materials Science and Engineering, Sweden, 2006.
[39] R.L. Smith. Maximum likelihood estimation in a class of nonregular cases. Biometrika, 72:67–90, 1985.
[40] M.A. Stephens. Edf statistics for goodness of fit and some comparisons. Journal of the American Statistical Association, 69:730–737, 1974.
[41] C.T. Volinsky and A.E. Raftery. Bayesian information criterion for censored survival models. Biometrics, 56:256–262, 2000.
[42] K. Wallin. Irradiation damage effects on the fracture toughness transition curve shape for reactor pressure vessel steels. Joint FEFG/ICF International Conference on Fracture of Engineering Materials and Structures, Singapore, 1991.
[43] K. Wallin. The Master Curve method: a new concept for brittle fracture. International Journal of Materials and Product Technology, 14:342–354, 1999.
[44] K. Wallin. Master curve analysis of the "euro" fracture toughness dataset. Engineering Fracture Mechanics, 69:451–481, 2002.
[45] K. Wallin. Objective comparison of the unified curve and master curve methods. International Journal of Pressure Vessels and Piping, 122:31–40, 2014.
[46] K. Wallin, A. Bannister, and P. Nevasmaa. New unified fracture toughness estimation scheme for structural integrity assessment. SINTAP Background Document, FITNET, 1999.
[47] K. Wallin and P. Nevasmaa. Structural integrity assessment procedures for european industry (sintap). Sub-Task 3.2 Report: Methodology for the Treatment of Fracture Toughness Data - Procedure and Validation. Report No. VAL A: SINTAP/VTT/7.VTT Manufacturing Technology, Espoo, 1998.
[48] Cochran W.G. The χ^2 test of goodness of fit. Annals of Mathematical Statistics, 23:315–345, 1952.
[49] F. Wilcoxon. Individual comparisons by ranking methods. Biometrics Bulletin, 1:80–83, 1945.
[50] S.H. Zanakis and J. Kyparisis. A review of maximum likelihood estimation methods for the three-parameter weibull distribution. Journal of Statistical Computation and Simulation, 29:419–428, 1986.
[51] X.-H. Zhu and J.A. Joyce. Review of fracture toughness (g,k,j,ctod,ctoa) testing and standardization. Engineering Fracture Mechanics, 85:1–46, 2012.

APPENDIX

A. SIMULATION OF EXPERIMENTAL DATASETS

A particular attention was paid to check that each of the 30 sets simulated in § 5.1 reproduce accurately the main features of the real EURO database. These data were simulated in compliance with a criterion of observed "distance" between regular and censored data in the original dataset. Let us consider a censored value $k_1(T_1)$ in the original dataset. For a small relative difference $\Delta_{T_2} = T_2/T_1 - 1$ between test temperatures, a small relative difference in the toughness value $\Delta_{K(T)} = K(T)/k_1(T_1) - 1$ and the small relative difference in the size of the test piece $\Delta_{B_T} = B(T)/b_1(T_1) - 1$, we define as a valid data the nearest value $k_2(T_2)$ to the direction of the experimental conditions by means of the following least weighted squares criterion:

$$k_2(T_2) \quad = \quad \arg\min_{k(T)} \left\{ \omega_1 \Delta_T^2 + \omega_2 \Delta_{k(T)}^2 + \omega_3 \Delta_{B_T}^2 \right\} \tag{9}$$

where $\{\omega_1, \omega_2, \omega_3\}$ are positive weights summing to 1. A Wilcoxon homogeneity statistical test [49] based on the closeness of the samples simulated in this way and the original sample showed that the first two terms of the above criterion play a major part in selecting a "good" value $k_2(T_2)$. The following choices were applied:

$$\{\omega_1, \omega_2, \omega_3\} \quad = \quad \{0.6, 0.3, 0.1\}.$$

The second test, founded on the two first criteria only, specifies a unique solution to the problem (9). The toughness dispersion

$$\Gamma_{k_1(T_1)} \quad = \quad k_1(T_1) - k_2(T_2)$$

is used to simulate a censored value $\tilde{k}_2(T_2)$ from a simulated valid value $\tilde{k}_2(T_1)$:

$$\tilde{k}_2(T_2) \quad = \quad \tilde{k}_2(T_1) + \Gamma_{k_1(T_1)}.$$

This bootstrapping procedure on the dispersion of toughness values generates datasets that are extremely similar to the original, as shown on Figures 9 and 10. Kolmogorov-Smirnov and Cramer-von Mises tests between truly observed and simulated data were finally conducted conditionally to several temperature values, which did not exhibit surprising behaviors.

[Figure 6 about here.]

[Figure 7 about here.]

B. TESTING THE RELEVANCE OF THE MASTER CURVE

Using the notations defined in (1) and (2), consider the three following toughness models differing by their degree of freedom (θ_i defining the vectors of unknown parameters to

estimate) and possible inequality constraints:

$$\begin{aligned}
\text{original Master Curve 1 (MC1):} &\quad \alpha = 4,\ K_{\min} = 20,\ \theta_1 = (b_1, b_2, b_3) \in \mathbb{R}_*^+;\\
\text{relaxed MC 2 (MC2):} &\quad \alpha = 4,\ \theta_2 = (K_{\min}, b_1, b_2, b_3) \in \mathbb{R}_*^+ \text{ and } K_{\min} > 20;\\
\text{relaxed MC 3 (MC3):} &\quad K_{\min} = 20,\ \theta_3 = (\alpha, b_1, b_2, b_3) \in \mathbb{R}_*^+ \text{ and } \alpha > 2;\\
\text{relaxed MC 4 (MC4):} &\quad \theta_4 = (\alpha, K_{\min}, b_1, b_2, b_3) \in \mathbb{R}_*^+ \text{ and } K_{\min} > 20 \text{ and } \alpha > 2.
\end{aligned}$$

Assume to dispose of the various MLE $(\hat{\theta}_i)_{i\in\{1:4\}}$ computed from the same dataset. Likelihood ratio tests (LRT) are one of the most powerful statistical tools [17] for conducting the following tests, that evaluate the statistical relevance of the original Master Curve :

(T1) H_0: MC1 versus H_1: MC2;

(T2) H_0: MC1 versus H_1: MC3;

(T3) H_0: MC1 versus H_1: MC4.

Such situations are instances of the general situation when the null hypothesis H_0 is defined by fixing r degrees of freedom of the encompassing model used in the alternative hypothesis (H_1): $r = 1$ in **(T1)** and **(T2)** and $r = 2$ in **(T3)**. Denoting $\mathcal{L}_i(\hat{\theta}_i)$ the likelihood of model MCi estimated in its MLE $\hat{\theta}_i$, the asymptotic distribution of the LRT statistic

$$R_{i,j} \quad = \quad 2\log\frac{\mathcal{L}_i(\hat{\theta}_i)}{\mathcal{L}_j(\hat{\theta}_j)}$$

is known under H_0. This distribution is dependent on r and the number of inequality constraints limiting the domain of definition of the test statistic. Since **(T1)** is equivalent to test if $K_{\min} = 20$ rather than $K_{\min} > 20$, then, assuming the MC hypothesis H_0 is true, the asymptotic distribution of $R_{1,2}$ is a mixture of Dirac δ_0 in 0 and chi-square distribution χ^2_r with $r = 1$ degree of freedom. More generally, based on Chapter 21 in [17], with n the number of regular observations:

$$\begin{aligned}
R_{1,2}, R_{1,3} &\overset{n\to\infty}{\sim} \frac{1}{2}\delta_0 + \frac{1}{2}\chi^2_1,\\
R_{1,4} &\overset{n\to\infty}{\sim} \frac{\alpha_{14}}{2\pi}\delta_0 + \frac{1}{2}\chi^2_1 + \left(\frac{1}{2} - \frac{\alpha_{14}}{2\pi}\right)\chi^2_2
\end{aligned}$$

where $\alpha_{14} = \cos\rho_{14}$ and ρ_{14} is the asymptotic linear correlation coefficient between α and $K_{\min}$, which can be consistently estimated using the correlation submatrix computed for $\hat{\theta}_4$. Note that an alternative to **(T1)** is simply to test if $K_{\min} = 20$ rather than $K_{\min} > 0$. In such a case, since $\mathcal{L}_2(\hat{\theta}_2)$ remains defined even if $K_{\min} \geq 0$, the Dirac term disappears and an usual χ^2_1 distribution is the asymptotic limit.

Additionally, testing *H_0: MC2 versus H_1: MC4* or testing *H_0: MC3 versus H_1: MC3* can be similarly conducted, as well as numerous other tests for more complicated functional forms given to α, $K_{\min}$ and $K_0 - K_{\min}$. However, it must be noticed that such tools cannot provide

a total-ordered testing strategy. Besides, numerous models are not embedded. Consequently, the LRT-based approach must be completed with other statistical testing procedures.

FIGURES

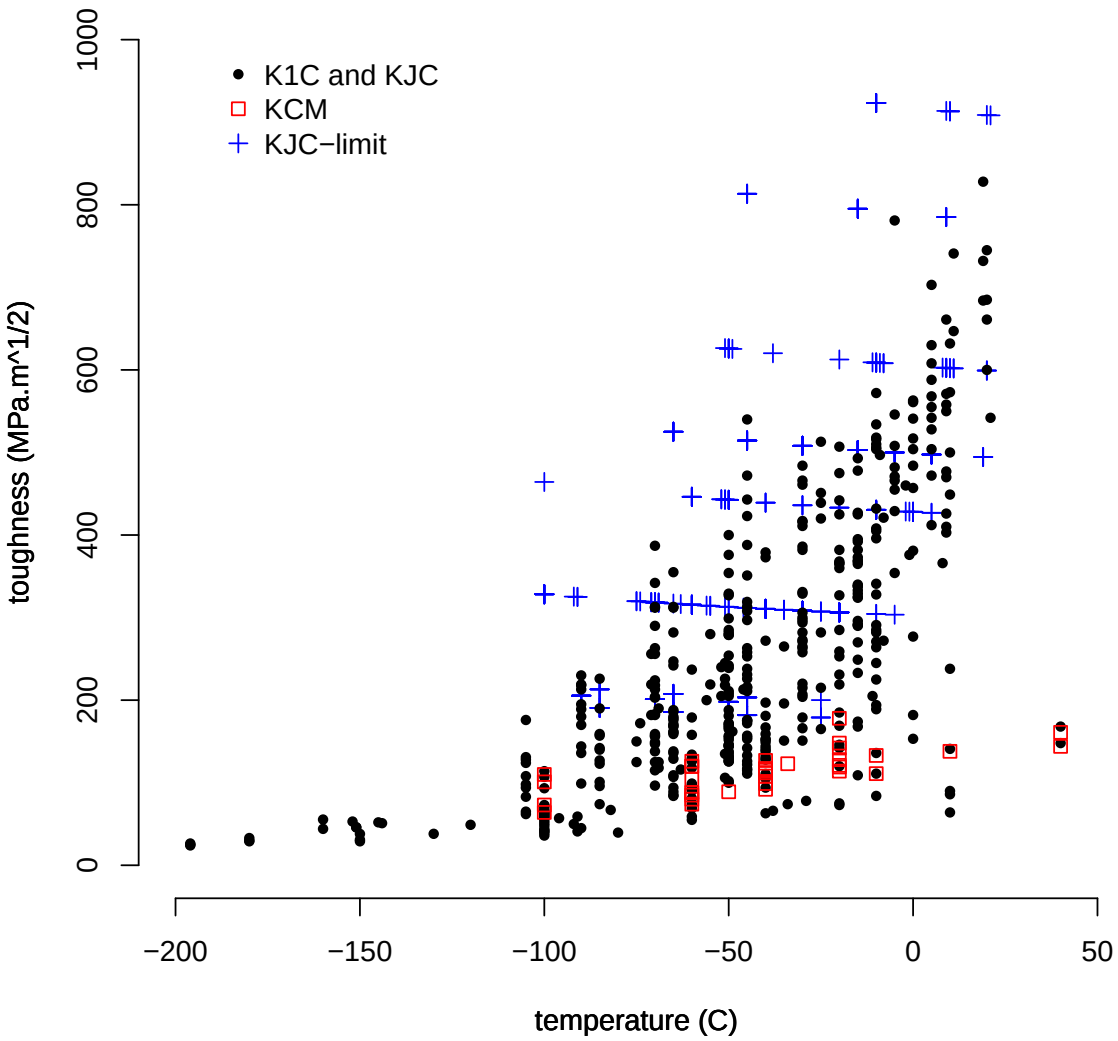

Figure 1. European fracture toughness database.

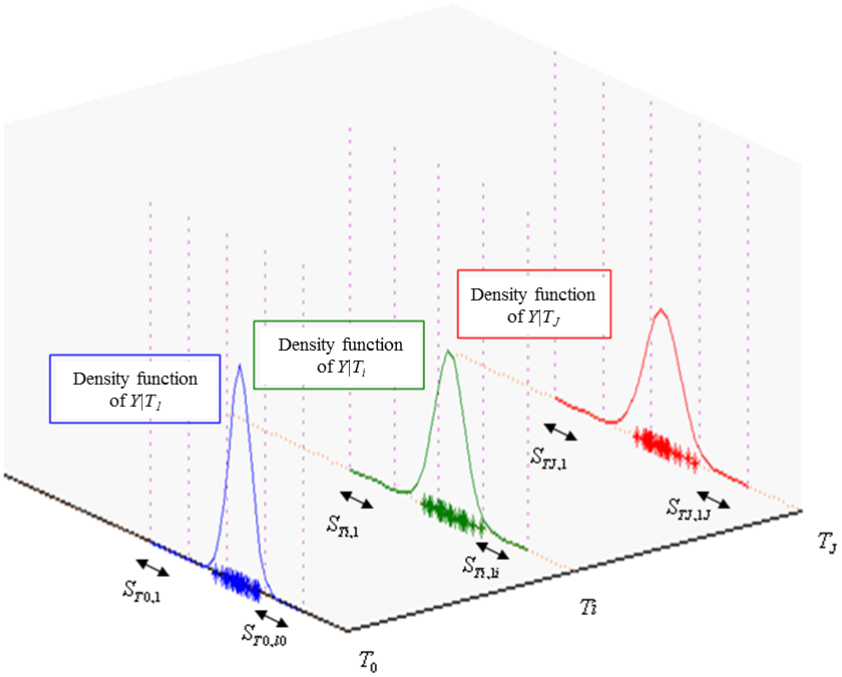

Figure 2. Subdividing the space of the variable $K_{IC}|T$.

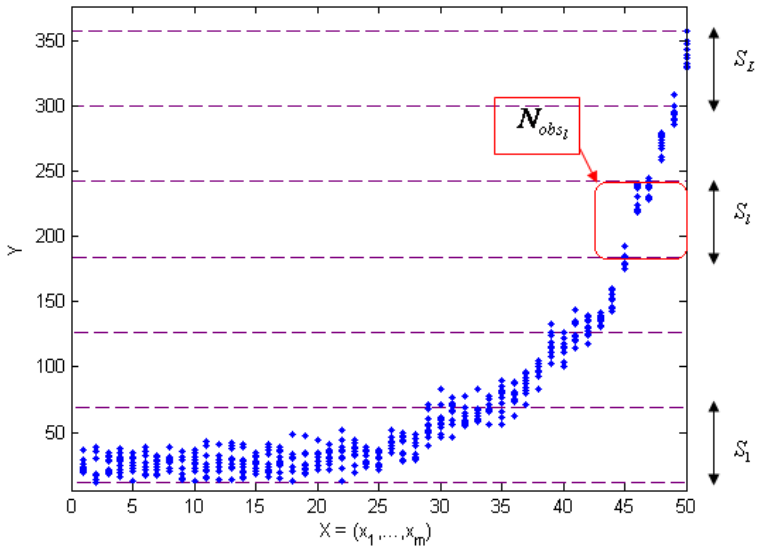

Figure 3. Subdivision of the space of a random variable $Y = K_{IC}$ dependent on a variable $X = T$.

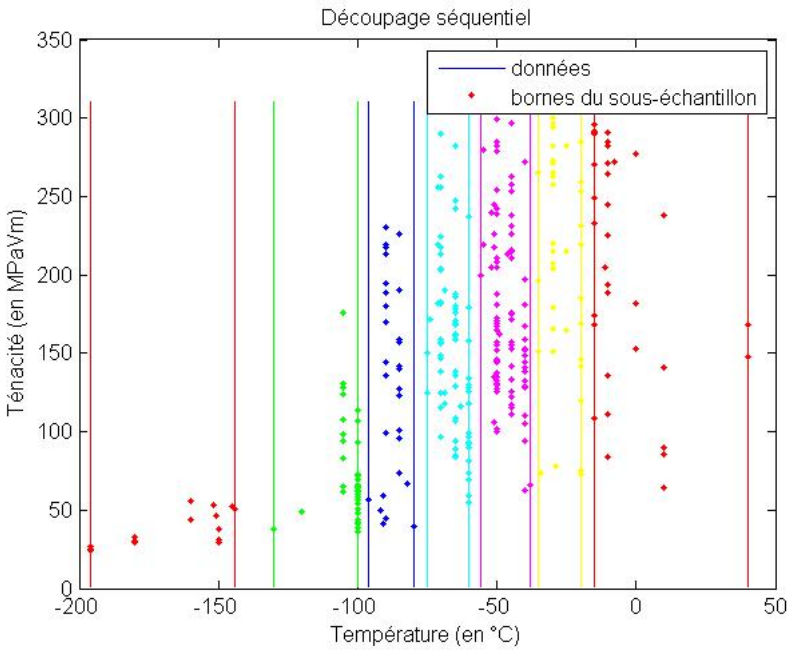

Figure 4. Sequential sub-sampling of the regular data from the EURO database.

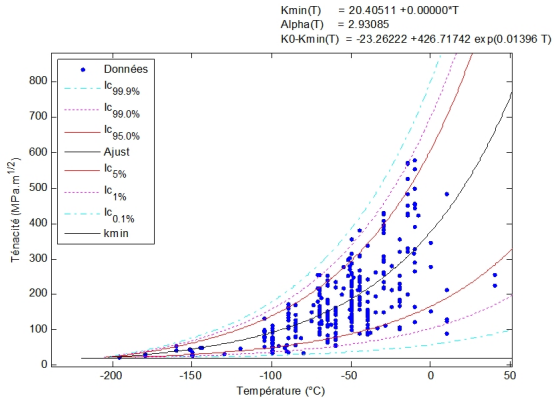

Figure 5. WOLF3 fitting on the European toughness database (valid data only).

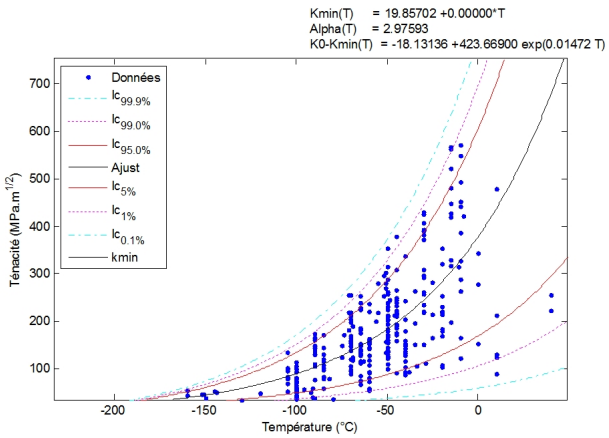

Figure 6. WOLF3 fitting on the European toughness database (all data).

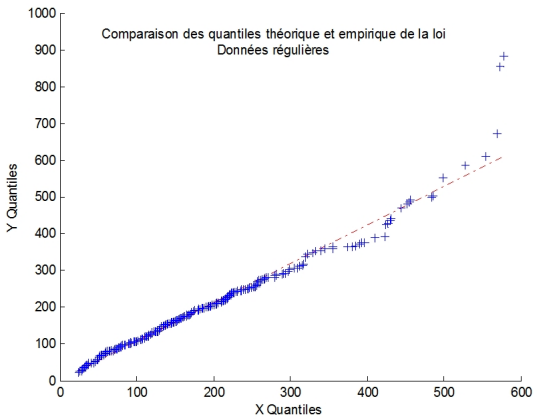

Figure 7. QQ plot of the WOLF3 fitting on the European toughness database (valid data only).

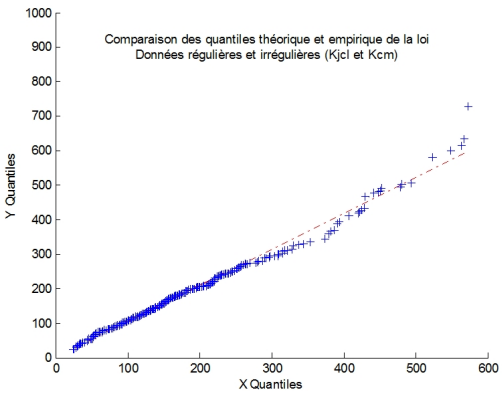

Figure 8. QQ plot of the WOLF3 fitting on the European toughness database (all data).

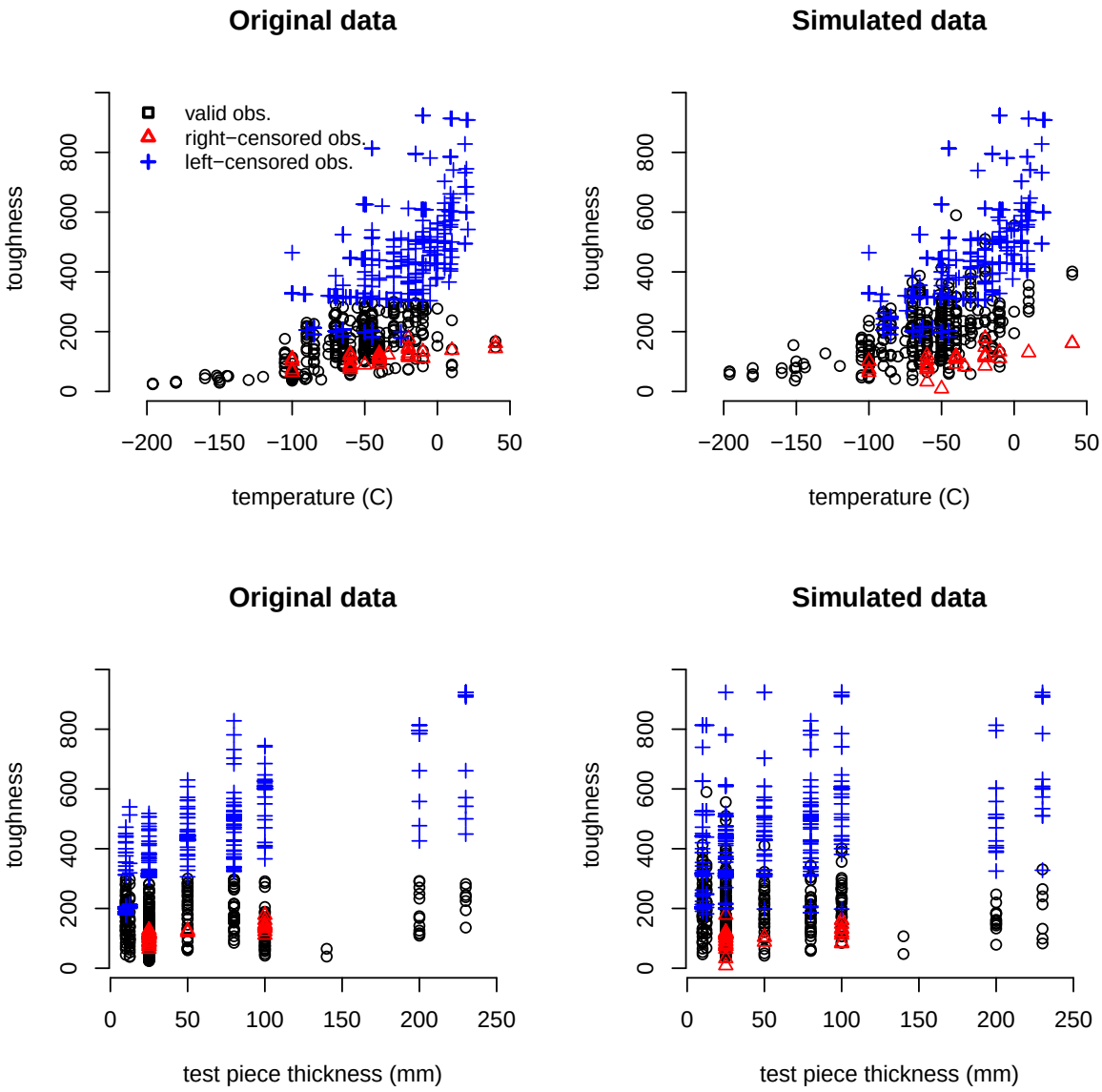

Figure 9. Example of a simulation of a toughness dataset from the EURO database (first experiment).

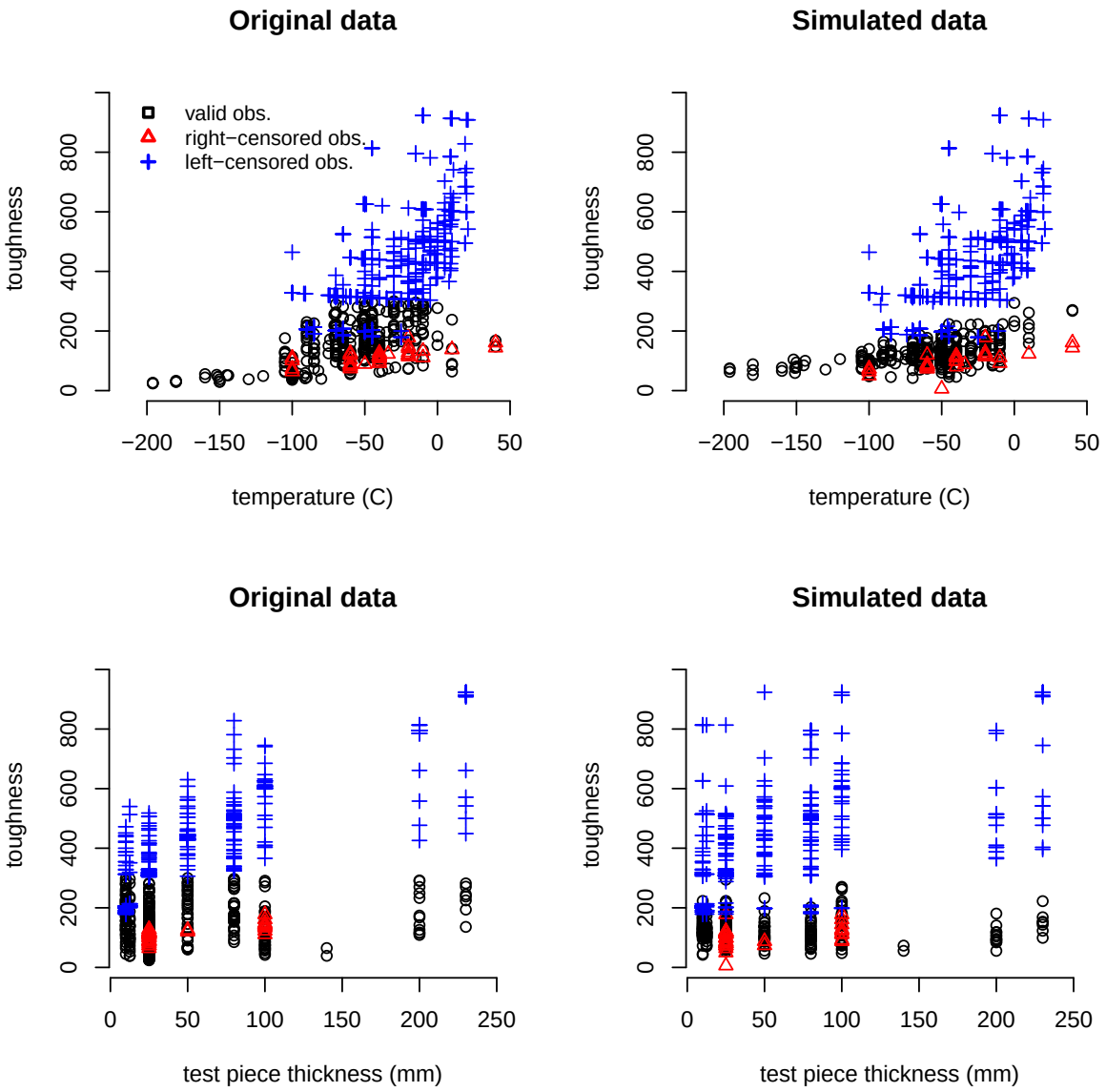

Figure 10. Example of a Master Curve simulation of a toughness dataset from the EURO database (second experiment).

TABLES

	Estimation based on regular data only	Estimation combining all types of data
α	2.78	2.90
$K_{\min}(T)$	$19.78 + 0.013 \cdot T$	$19.85 + 0.048 \cdot T$
$K_0(T) - K_{\min}(T)$	$0.0002 + 484.20 \cdot \exp(0.0186 \cdot T)$	$0.0012 + 463.23 \cdot \exp(0.0175 \cdot T)$
Graphic representation	Fig. 5	Fig. 6
QQ dispersion [0.1% - 99%]	2.30	2.16
QQ dispersion [75% - 99%]	5.81	5.30
QQ dispersion [0.1% - 20%]	2.62	2.57
QQ plot	Fig. 7	Fig. 8

Table 1. Results of estimation by the method of maximum likelihood on the EURO fracture toughness database.

	estimated mean	standard deviation	simulated value
α	3.141	0.12	3
$K_{\min}$ (original ordinate)	19.623	2.41	20
$K_{\min}$ (slope)	0.00730	0.0087	0
$K_0 - K_{\min}$ (original ordinate)	6.558	8.55	2
$K_0 - K_{\min}$ (slope)	423.065	15.38	424
$K_0 - K_{\min}$ (exponential coefficient)	0.001498	0.0012	0.001472

Table 2. Results of the estimates averaged over the 30 datasets simulated from estimates based on the EURO database.

	Choice of functionals	
	linear	shifted exponential
$\boldsymbol{\alpha}$	$\mathbf{K}_{\min}(\mathbf{T}) = a_1 + a_2 \cdot T$	$\mathbf{K_0}(\mathbf{T}) - \mathbf{K}_{\min}(\mathbf{T}) = b_1 + b_2 \exp(b_3 \cdot T)$
4.18 (0.19)	$a_1 = 20.55$ (3.12)	$b_1 = 52.90$ (8.4)
	$a_2 = 0.015$ (0.024)	$b_2 = 196.21$ (11.6)
		$b_3 = 0.0203$ (0.003)
	shifted exponential	shifted exponential
$\boldsymbol{\alpha}$	$\mathbf{K}_{\min}(\mathbf{T}) = a_1 + a_2 \exp(a_3 \cdot T)$	$\mathbf{K_0}(\mathbf{T}) - \mathbf{K}_{\min}(\mathbf{T}) = b_1 + b_2 \exp(b_3 \cdot T)$
4.22 (0.28)	$a_1 = 14.20$ (6.37)	$b_1 = 53.26$ (9.67)
	$a_2 = 11.68$ (11.84)	$b_2 = 186.56$ (13.58)
	$a_3 = 0.0540$ (0.04)	$b_3 = 0.0188$ (0.0024)

Table 3. Average estimation results (standard deviation within parentheses) for the Master Curve model for simulated datasets.

Model indic.	$K_{\min}$	α	$K_0 - K_{\min}$	transition temp. T_0	AIC	χ^2 test $p-value$	Figur (SOM
MC	fixed (20)	fixed (4)	shifted exponential	-97.393°C	3545.93	1.10^{-6}	MC-1
1	linear	linear	linear	-99.99°C	3567	1.10^{-6}	S-1
2	linear	linear	quadratic	-86.47°C	3414	0.0043	S-3
3	linear	linear	shifted exponential	-91.43°C	3413	0.222	S-5
4	quadratic	linear	linear	-99.99°C	3462	1.10^{-6}	S-7
5	quadratic	linear	quadratic	-93.05°C	3410	0.0914	S-9
6	quadratic	linear	shifted exponential	-89.74°C	3414	0.4638	S-11
7	shifted exponential	linear	linear	-99.99°C	3544	1.10^{-6}	S-13
8	shifted exponential	linear	quadratic	-88.69°C	3388	0.0305	S-15
9	shifted exponential	linear	shifted exponential	-90.65°C	3414	0.4859	S-17
10	linear	constant	linear	-99.99°C	3613	1.10^{-6}	S-19
11	linear	constant	quadratic	-90.58°C	3420	0.232	S-21
12	linear	constant	shifted exponential	-89°C	3415	0.5751	S-23
13	quadratic	constant	linear	-99.99°C	3603	1.10^{-6}	S-25
14	quadratic	constant	shifted exponential	-91.04°C	3412	0.5975	S-27
15	shifted exponential	constant	linear	-99.99°C	3595	1.10^{-6}	S-29
16	shifted exponential	constant	quadratic	-88.70°C	3428	0.1438	S-31
17	shifted exponential	constant	shifted exponential	-89.91°C	3412	0.4328	S-33

Table 4. Model comparisons for the EURO database (valid toughness data only). Figures MC-1 and S-i refer to figures presented within the Supplementary Online Material (SOM) that accompanies this article. MC is for the usual Master Curve.

Model indic.	$K_{\min}$	α	$K_0 - K_{\min}$	transition temp. T_0	AIC	χ^2 test $p-value$	Figure (SOM)
1	linear	linear	linear	-99.99°C	3601	1.10^{-6}	S-2
2	linear	linear	quadratic	-86.26°C	3461	0.0044	S-4
3	linear	linear	shifted exponential	-89.17°C	3460	0.687	S-6
4	quadratic	linear	linear	-99.99°C	3579	1.10^{-6}	S-8
5	quadratic	linear	quadratic	-86.14°C	3447	0.0615	S-10
6	quadratic	linear	shifted exponential	-92.65°C	3461	0.4351	S-12
7	shifted exponential	linear	linear	-99.99°C	3579	1.10^{-6}	S-14
8	shifted exponential	linear	quadratic	-89.26°C	3467	0.0164	S-16
9	shifted exponential	linear	shifted exponential	-91.03°C	3466	0.3945	S-18
10	linear	constant	linear	-99.99°C	3651	1.10^{-6}	S-20
11	linear	constant	quadratic	-93.61°C	3460	0.0565	S-22
12	linear	constant	shifted exponential	-90.4°C	3466	0.85	S-24
13	quadratic	constant	linear	-99.99°C	3711	1.10^{-6}	S-26
14	quadratic	constant	shifted exponential	-91.52°C	3461	0.7191	S-28
15	shifted exponential	constant	linear	-99.99°C	3685	1.10^{-6}	S-30
16	shifted exponential	constant	quadratic	-86.57°C	3523	0.0708	S-32
17	shifted exponential	constant	shifted exponential	-91.63°C	3461	0.8113	S-34

Table 5. Model comparisons for the EURO database (all toughness data). Figures S-X refer to figures presented within the Supplementary Online Material (SOM) that accompanies this article. The χ^2 tests are conducted only by confronting assessed models with the uncensored empirical distribution.