

Estimating a flaw size distribution using data from both destructive tests and non-destructive in-service inspections

Merlin Keller^{a*}, Nicolas Bousquet^a,

^aEDF Research & Development

Department of Industrial Risk Management, Chatou, France

Abstract: Among the stochastic inputs entering into the probabilistic calculation of the reliability of industrial equipments/components of electric power plants, the flaw size is often a key variable whose distribution has to be carefully estimated. Although the considered component flaws appear as manufacturing defects and remain non-evolutive over time, the assessment of this distribution is usually not trivial. Indeed, the data that may be used to this purpose typically come from nondestructive experiments (NDE), affected by observational noise and progressive censoring due to the detection limits of the testing process.

In this paper, we show how a combination of these data with observations coming from destructive experiments allows to estimate the flaw size distribution, using maximum likelihood techniques, as well as investigate the features of the probability of detection (POD) function.

Keywords: destructive and non-destructive testings, probability of detection, noisy experiments

1. INTRODUCTION

Risk studies on highly reliable industrial components are often limited by the very low number of observed failures. Alternatively, the probability of a failure can be computed through numerical experiments involving computer codes reproducing the physical (e.g., thermomechanical) behavior of the studied component in its working environment. In order to explore exhaustively the configurations of input values of such numerical codes leading to failure scenarios, these inputs are typically considered as stochastic. In many structural reliability studies, a key input is the distribution of flaw sizes. For certain passive components, such flaws typically appear as manufacturing defects that remain non-evolutive with time and whose size can be summarized by their height and width.

Assessing the distribution of such flaw sizes is usually done by solving a probabilistic inverse problem, associated with indirect noisy measurements from non-destructive in-service inspections. In the present case, the inspections were led by ultrasonic devices. These inspections are characterized by an observational noise and a probability of detection (POD) related to the performance of the examination process, depending mainly of the limiting flaw size (its height). In the problem considered here, a second source of data is available: destructive testing produces data which are statistically considered as direct measures of flaw sizes.

In this article, we establish a general framework for the statistical estimation by maximum likelihood of the flaw size distribution, while accounting for both observational noise and POD, as pre-specified by the user. This framework is based on a statistical model, which we describe in Section 2. Numerical integration and asymptotic statistics are then used to provide estimators and confidence regions, respectively, as described in Sections 3. These are based on a user-specified POD, usually incorporating expert judgement, as explained in Section 4. Additionally, checking for the relevance of the chosen POD can easily be done within our framework using the two sources of data. This approach is demonstrated on simulated data in Section 5, and results obtained on real data are presented in Section 6, which highlight the important role played by the POD function when estimating flaw size distribution from NDE data. We conclude in Section 7 by a brief discussion on the perspectives opened by this work.

2. MODELING

Denote X the random variable associated with a flaw height. The width of a flaw appears significantly larger than its height, so that it can be considered as not affected by observational noise. Therefore we mainly focus on the marginal distribution of X in this paper. Previous studies, led on data coming from destructive experiments, have shown that a relevant modeling for X is the Weibull distribution with probability density function (pdf):

$$f(x|\eta, \beta) = \beta/\eta (x/\eta)^{\beta-1} \exp(-(x/\eta)^\beta) \mathbf{1}_{\{x>0\}}.$$

The parameter vector $\theta = (\eta, \beta)$ must also be estimated from the data. Denote Z the random variable associated with the observation of X through an inspection process, and assume to have noisy observations $\mathbf{z} = (z_1, \dots, z_n)$ such that $z_i = x_i \cdot e_i$ where, whatever $i = 1, \dots, n$, the multiplicative observational error e_i follows a known distribution with pdf $f_\epsilon(\cdot)$. If the inspection were *destructive* (testing), the observational noise could be considered negligible (ie., $e_i \equiv 1$). For *non-destructive* inspections, e_i follows a Gamma distribution centered on 1 (or, nearly equivalently, a truncated Gaussian distribution), with a fixed coefficient of variation ρ . Information about ρ is typically provided by manufacturer specifications and performance testings of ultrasonic processes. The choice of a multiplicative noise is explained by the positivity of all quantities involved.

Furthermore, it is assumed there exists a unknown number m of missing measures $(z_{n+1}, \dots, z_{n+m})$, *upperly censored* by the detection limits of the measurement process, so that the full vector of latent variables is $\mathbf{x} = (x_1, \dots, x_{n+m})$. Denote then D the binary random variable indicating whether X was detected or not. It depends on the value of X through a function defined in $[0, 1]$ and known as the *probability of detection* (POD):

$$\text{POD}(x) = P(D = 1|X = x).$$

In the following, $\text{POD}(x)$ is assumed to be an increasing known function of x . The unconditional POD is given by

$$P_f(\theta) = P(D = 1|\theta) = \int_{x \in \mathbb{R}} \text{POD}(x) f(x|\theta) dx. \quad (1)$$

Besides, from Celeux et al. [1], given the number of observation n and θ , m follows a negative binomial distribution $\mathcal{NB}(n, 1 - P_f(\theta))$:

$$P(M = m|\theta) = \frac{(n + m - 1)!}{(n - 1)! m!} P_f(\theta)^n (1 - P_f(\theta))^m. \quad (2)$$

The vector of parameters θ may then be estimated by maximizing the likelihood of observations. If m, \mathbf{z} and \mathbf{x} were observed, the complete likelihood of the model would be:

$$\begin{aligned} \ell(m, \mathbf{z}, \mathbf{x}|\theta) &= P(M = m|\theta) p(\mathbf{x}|\theta, m) p(\mathbf{z}|\mathbf{x}) \\ &= \frac{(m - 1)!}{(n - 1)! (m - n)!} P_f(\theta)^n (1 - P_f(\theta))^{m-n} \\ &\times \prod_{i=1}^n \frac{\text{POD}(x_i) f(x_i|\theta) f_\epsilon(z_i/x_i)/x_i}{P_f(\theta)} \times \prod_{i=n+1}^m \frac{(1 - \text{POD}(x_i)) f(x_i|\theta)}{1 - P_f(\theta)}. \end{aligned} \quad (3)$$

Therefore the likelihood of the truly observed data (in-service inspections) is obtained by integrating out the latent variables (M, X) from the above complete likelihood, and can easily be seen to be equal to:

$$\ell(\mathbf{z}|\theta) = \prod_{i=1}^n P_f^{-1}(\theta) \int \text{POD}(x_i) f(x_i|\theta) f_\epsilon(z_i/x_i)/x_i dx_i. \quad (4)$$

3. MAXIMUM LIKELIHOOD MAXIMIZATION

Preparatory works showed that the likelihood of a mixture of noisy Weibull data and statistically perfect data (coming from destructive testing) can suffer from identifiability issues when the progressive censoring characterizing the observational process is not accounted for, as testified on Figure 1. It was found that when the all data suffer from noise measurement (typically $\rho = 20\%$), those identifiability issues become worse. Nevertheless, accounting for both effects, the likelihood (4) can be maximized. Since no explicit conditional solution can emerge from likelihood equations, numerical optimization routines must be used. Tests demonstrated that a Gaussian quadrature method [2] can yield good approximations of the integrals involved in the above likelihood, and can be successfully coupled with a standard optimization method as Nelder-Mead [3] which furthermore avoids to specify the Jacobian vector $\partial\ell(\mathbf{z}|\theta)/\partial\theta$, as required by typical descent gradient methods (e.g., Newton-Raphson algorithms).

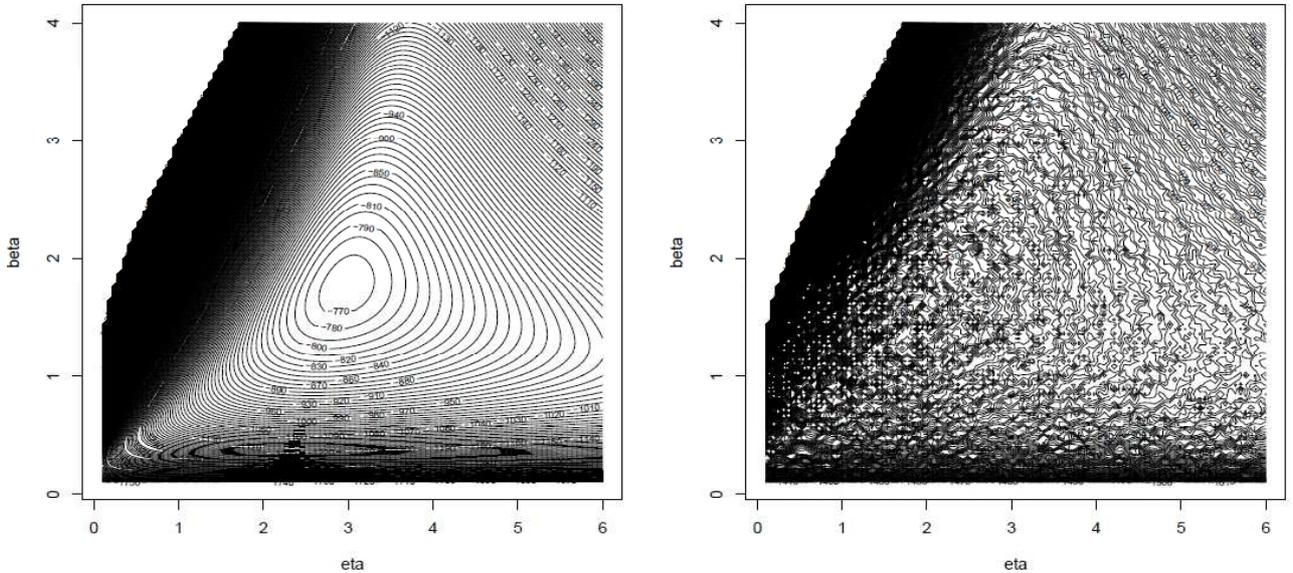


Figure 1: Level lines of the Weibull loglikelihood of $n = 430$ simulated data ($\beta = 1.8$, $\eta = 3.09$). *Left:* data are sampled without perturbation by observational noise, leading to an identifiable maxima. *Right:* 64% of the sampled data are affected by noise measurement ($\rho = 20\%$), which implies a strong loss of regularity implying identifiability issues of numerical optimization routines. The ratio of affected data is close to that of the real industrial data considered in this paper.

Nonetheless, such methods were found to be time-consuming, especially so when combined with bootstrap methods to estimating the variability of the parameter estimators and the functions of interest. This is especially the case when considering the probability $p(h)$ for a flaw height, observed through the noisy experimental process, to exceed the value h . In the experiments described in the following sections, such bootstrap methods took nearly 66 hours on an AMD Sempron 3600. Denoting $\hat{p}(h)$ the MLE of $p(h)$, an alternative Delta method can be carried out [4], implying:

$$\text{Var}[\hat{p}(h)] \approx \nabla F(h|\hat{\theta})^\top \hat{\Sigma} \nabla F(h|\hat{\theta})$$

when $\hat{\theta} = (\hat{\eta}, \hat{\beta})$ denotes the maximum likelihood estimator (MLE), $\hat{\Sigma}$ is the inverse of the estimated Fisher information and F is the cumulative distribution function (cdf)

$$F(h|\theta) = \frac{\int \text{POD}(x) f(x|\theta) F_\epsilon(h/x) dx}{P_f(\theta)} \quad (5)$$

with $F_\epsilon(h/x) = \int_0^h f_\epsilon(u/x) du$. The derivatives involved in the calculus of $\text{Var}[\hat{p}(h)]$ (providing confidence intervals around $p(h)$) can easily be computed using numerical integration.

4. CALIBRATING THE PROBABILITY OF DETECTION (POD) THROUGH EXPERT JUDGMENT

The POD is the most common indicator for quantifying the reliability and sensitivity of non-destructive in-service inspections [5, 6]. Its calibration appears as one of the most important methodological issues encountered in the area of performance testings, especially in nuclear contexts [7]. Such calibrations are based on a set of experimental performance measures, that are more and more often complemented with numerical simulations reproducing the measurement process itself [8], especially to improve the validation and certification of these processes in industrial contexts or/and the training of manipulators [9, 10]. In this paper, we focalize on the performance measures. They can be either destructive, requiring costly ground samples, or non-destructive when benefiting from mock-ups that are machined with artificial flaws. In standard testing conditions, the certification of the ultrasonic process allowed to establish the precise limits of non-detection (POD=0%) and full detection (POD=100%), and to carry out a *signal response* method [11, 12] ; the logarithms of X and the signal S are linearly linked with normally-distributed random residuals ϵ' :

$$\log S = \beta_0 + \beta_1 \log X + \epsilon',$$

which leads to the parametric form promoted by Berens [12]:

$$\text{POD}(x) = \Phi\left(\frac{\log(x) - \log(x_m^*)}{1/q}\right) \quad (6)$$

where x_m^* is the height for which 50% of defects are detected, and q measures the quality of the POD (directing the average slope of the detection curve). However, the non-evolutive feature of the defects considered here does not allow to establish precise POD curves that are representative of the ground behavior as it could be done, for instance, when dealing with corrosion effects [13] inducing that successive measures can yield information. Therefore expert knowledge was introduced into the modeling to account for undesirable effects (e.g., due to false alarms) and form constraints (convexity properties of the POD).

5. SIMULATION EXPERIMENTS

In the experiments summarized in this section, the simulated data were produced as follows:

- $m_E = 190$ data assimilated to *destructive measures* follow the Weibull distribution $\mathcal{W}(\eta_0 = 3.09, \beta_0 = 1.8)$;
- $m_{IS} = 340$ data assimilated to *non-destructive measures* are produced using the same distribution but are submitted to a observational multiplicative noise with coefficient of variation $\rho = 20\%$ and the probability of detection (6) calibrated with $x_m = 6$ and $q = 20$ (this POD is plotted on Figure 4). This simulation can be easily produced by the following accept-reject algorithm: for $i = 1, \dots, m_{IS}$,
 1. **sample** $x_i \sim f_W(\cdot | \eta_0, \beta_0)$ and $u_i \sim \mathcal{U}_{unif}[0, 1]$;
 2. **produce** $z_i = x_i \cdot \epsilon_i$ with $\epsilon_i \sim f_\epsilon$;
 3. **accept** z_i if $u_i \leq \int_{\mathbb{R}^+} \text{POD}(z_i/u) f_\epsilon(u) du$.

It must be noticed that the phenomenon of detection produces non-destructive size measures that are significantly higher than the destructive measures, so that the former usually provide a truncated image of the common distribution through its upper tail. An example of sampled data and estimated distributions are plotted on Figure 2. The MLE of (η, β) and $P_f(\theta)$ based on the mixture of destructive and non-destructive measures appeared to be a good estimator of these quantities, as well as the confidence intervals which were produced, as illustrated on Figure 3.

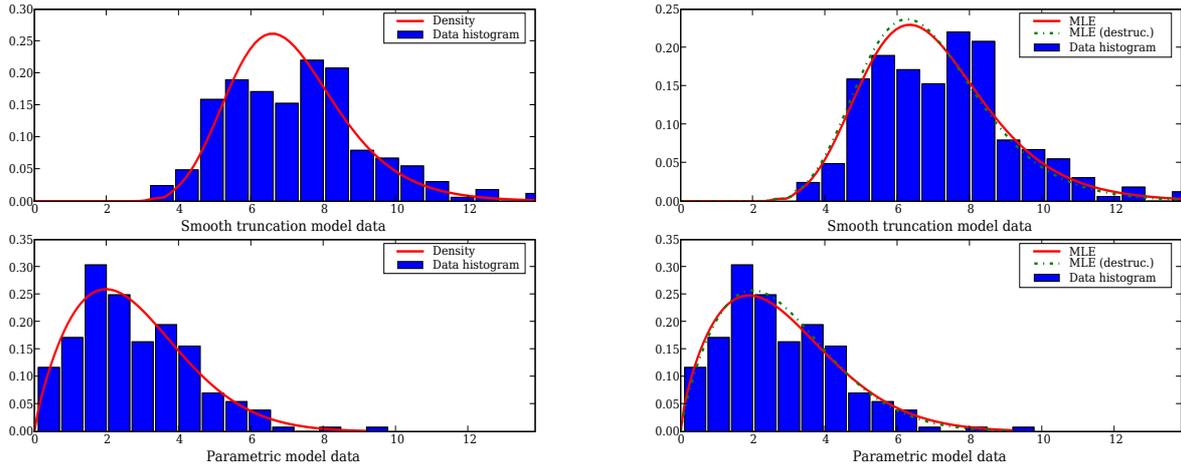


Figure 2: Histograms of simulated data. *Top*: non-destructive test data. *Down*: destructive test data. *Left*: true density function (solid line). *Right*: density function estimated by maximum likelihood on the basis of the complete data (MLE) or the destructive test data only (MLE (destruc.))

The variations between the MLE results based on the mixture of destructive and non-destructive data and the MLE based on the destructive data alone were always found to be statistically negligible. This feature allows to foresee the possibility of checking and strengthening the weakest aspects of the modeling, namely the parametric model chosen for the POD function.

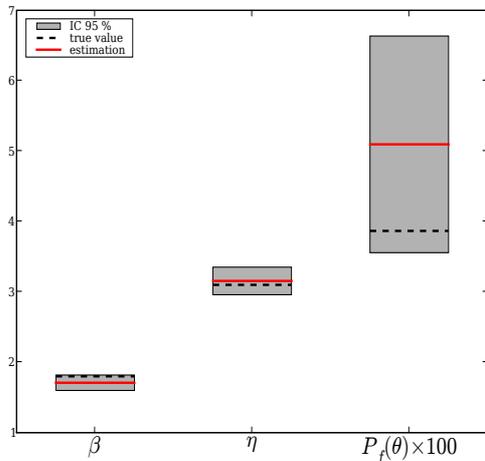


Figure 3: Estimates and 95% confidence intervals on (η, β) and $P_f(\theta)$, based on a mixture of non-destructive and destructive test data.

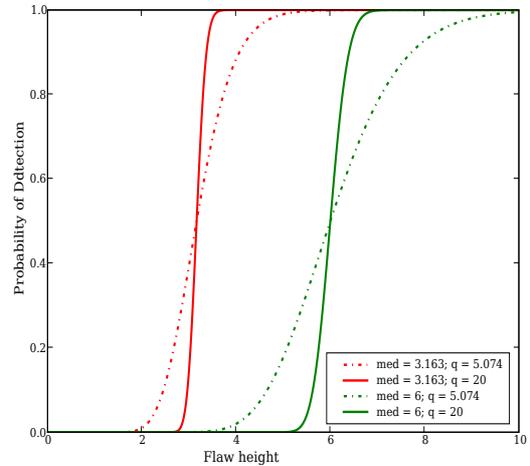


Figure 4: Two families of probability of detection. “med” is for the median fractile x_m^* in Equation (6).

6. APPLICATION TO REAL DATA: CHECKING THE RELEVANCE OF THE POD

The real available data are composed with 190 flaws observed on fractions of a steel component submitted to destructive laboratory experiments, and 341 couples of defect size (height and length) produced by ground experiments involving a ultrasonic device. Prior knowledge about the feature of this device was available to establish a series of four PODs (Figure 4). The differences between the PODs resulted from considerations about extreme behaviors (idealistic and pessimistic ground

conditions) and variations of the state of the component surface. Supplementary information coming from the certification of the device was not used to establish these PODs, but was kept to offer a critical view of the estimation results.

The estimated cumulative distribution function of noisily observed flaw sizes, for each POD, are plotted on Figure 5. It can be seen that the POD calibrated with $x_m^* = 6$ and $q = 5.074$ provides the better fit with the empirical distribution function of the noisily observed flaw size, after a likelihood maximization involving the two databases. This POD indeed is the most in accordance with the technical knowledge arising from experiments on mock-ups. Especially, in the converted scale used in this article, the theoretical complete detection limit (in standard experimental conditions) was certified at the value $x = 8.5$.

However, even when using this ‘best-fitting’ POD, one can see that the confidence intervals are too small to include the empirical data CDF, suggesting that all sources of uncertainty are not yet properly accounted for in the model used here. On one hand, it may be the case that the observation noise is under-estimated. On the other hand, it is obvious from the empirical cdf’s in Figure 4 that the NDE data is rounded, a fact totally unacknowledged in our model and inherent to the ultrasonic device used to produce these measures. We have re-computed the estimated cdf’s accounting for this rounding effect (not shown here), however even in this case we did not obtain a satisfactory matching between the proposed model and the data.

7. CONCLUSIONS AND PROSPECTIVE AVENUES

In this paper we have introduced a general methodology for maximum likelihood estimation of the flaw size distribution of a passive industrial component, using either:

- data from non-destructive in service inspection, such as ultrasonic devices, plagued by measurement noise and progressive censoring, summarized through the POD function,
- data from destructive experiments, which can be considered as a perfect sample from the target distribution,
- or a combination of both types of data.

Our methodology requires the user to specify a measurement noise distribution and a POD function. Our simulation results show that, provided that correct values are given for these quantities, precise estimation of the flaw size distribution is possible. Furthermore, these results also suggest that the first type of data may be superfluous when sufficiently many destructive test data are available. In such cases, non-destructive inspection data are essentially useful to verify that the user-specified POD function is indeed in agreement with the data.

When applied to a real dataset, this approach showed that the distribution for the non-destructive inspection data predicted based on a POD based expert judgment was in fact significantly different from the real data distribution. Note that this is not a major issue if estimating the flaw size distribution is the main goal of the analysis, since this can be done in this particular case directly from the destructive test data. However, in other cases only non-destructive inspection data are available, hence a precise knowledge of the POD is required. For this reason, we consider that the work presented here must be continued along the following directions:

- To begin with, analyzing the influence of the flaw size distribution on the final result of the reliability study, that is, the probability of failure of the studied component, is important in order to specify what precision is awaited in the estimation. Likewise, assessing the number of

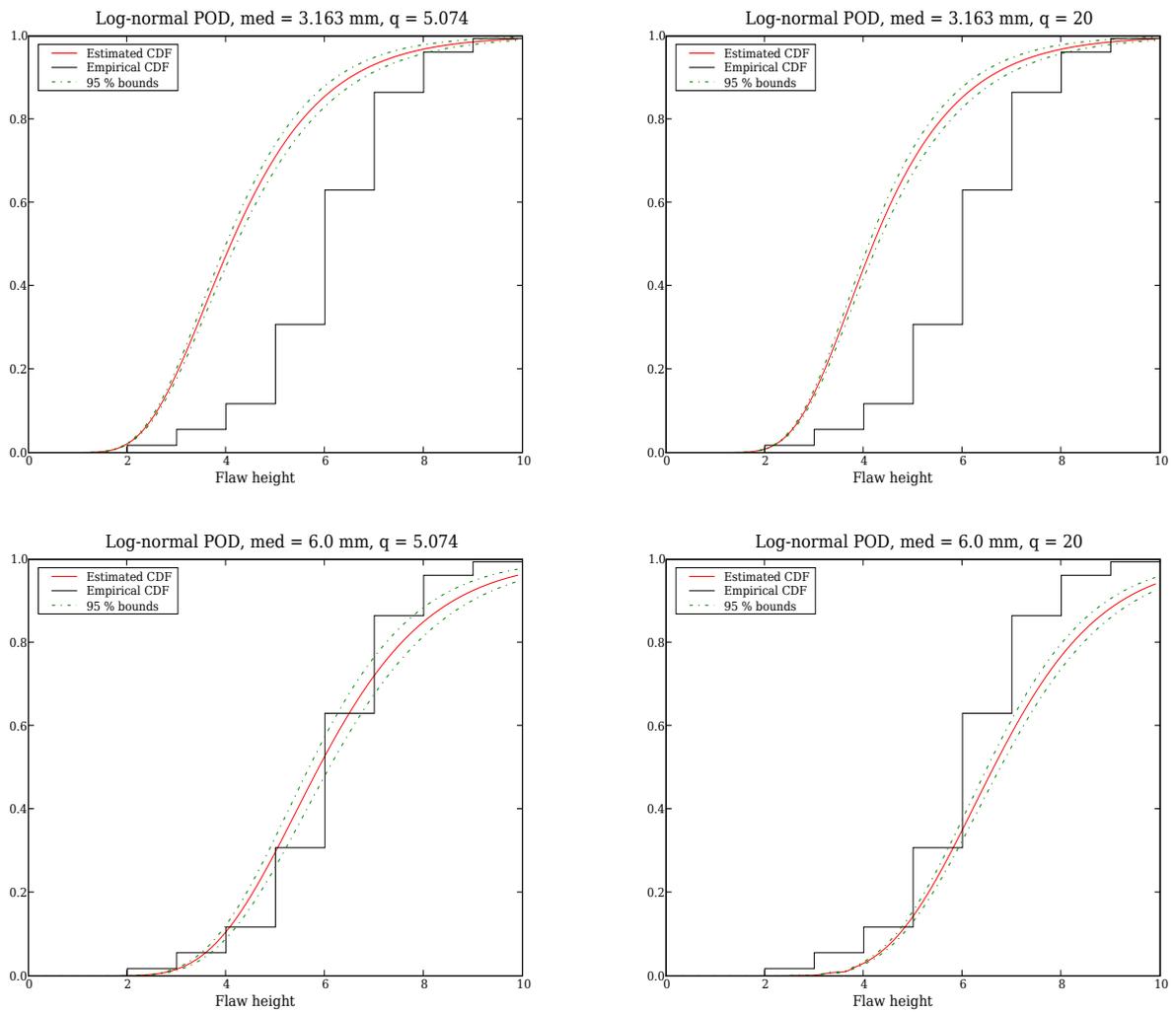


Figure 5: Estimation of the cumulative distribution function (cdf) of noisily observed flaw sizes as a function of the choice of POD, compared to their empirical cdf.

observations from non-destructive inspection and/or destructive tests necessary to attain such a precision can help plan future in service inspection strategies as well as destructive tests experiments. Both goals can be done through a sensitivity analysis similar to the (simplistic) one done here to study the influence of the POD function on the non-destructive inspection data distribution estimation.

- Furthermore, developing an algorithm for maximum likelihood estimation of the POD itself, along with the associated confidence intervals, can be done in our framework using the two available sources of data. Such a data-driven POD could then be used in future studies, for instance to calibrate a computer code emulating the non-destructive measurement process.
- Finally, based on the previous steps, both the POD and the flaw size distribution could then be estimated jointly by combining both sources of data. This could be useful if only few destructive tests have been conducted, and if the precise form of the POD function is not precisely known. Such an estimation would probably best be done in a Bayesian framework, which would allow to incorporate expert opinion, *e.g.* on the POD, under the form of a prior distribution, and more generally exploit efficiently all available sources of information.

REFERENCES

- [1] Celeux G., Persoz M., Ngatchou-Wandji J., and Perrot F. Bayesian modelling of PWR Vessels flaw distributions. *Reliability Engineering and System Safety*, 66:243–252, 1999.
- [2] P. J. Davis and P. Rabinowitz. *Methods of numerical integration*. Academic Press, New York, 1975.
- [3] Nelder J. and Mead R. A simplex method for function minimization. *The Computer Journal*, 7:308–318, 1965.
- [4] A.W. van der Vaart. *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2000.
- [5] Forsyth D.S., Safisadeh M.S., and Fahr A. Issues in the determination of probability of detection using field inspection data. *Proceedings of the 3rd European-American Workshop on Reliability of NDE and Demining*, 2002.
- [6] Simola K., Cronvall O., Männistö I., Gunnars J., Alverlind L., Dillström P., and Gandossi L. Studies on the effect of flaw detection probability assumptions on risk reduction at non-destructive inspection. *Proceedings of the ESREL 2010 Conference*, 2010.
- [7] EPRI. Reactor pressure vessel inspection reliability based on performance demonstrations. *EPRI Research Report, Palo Alto, CA:2004.1007984*, 2004.
- [8] Moles M.D.C. Orchid - a computer simulation of the reliability of an nde inspection system. *Journal of Nondestructive Evaluation*, 6:23–31, 2004.
- [9] Lhemery A., Calmon P., and Paradis L. Ultrasonic field simulations for scan coverage studies. *Review of Progress in QNDE. D.O. Thompson and D.E. Chimenti, Eds.*, 1998.
- [10] Leymarie N., Calmon P., Fouquet T., and Schumm A. Semi-analytical-fem hybrid modeling of ultrasonic defect responses. *Proceedings of the ECNDT 2006 Conference*, 2006.
- [11] Harris D.O. A means of assessing the effects of nde on the reliability of cyclically loaded structures. *Material Evaluation*, 35:57–65, 1977.
- [12] Berens A.P. Nde Reliability Data Analysis. *Metals Handbook Vol.17. Nondestructive Evaluation and Quality Control (pp. 689-701)*, 1985.

- [13] Rudlin J.R. and Kenzie B.W. Probability of detection for corrosion defects. *Proceedings of the 3rd European-American Workshop on Reliability of NDE and Demining*, 2002.