Merging of Experimental and Simulated Data Sets with a Bayesian Technique in the Context of POD Curves Determination

Bastien CHAPUIS *, Nicolas DOMINGUEZ *, Frédéric JENSON *,
Pierre WILLAUME **, Thierry YALAMAS **
* CEA LIST Centre de Saclay, France
** PHIMECA Engineering SA, Cournon d'Auvergne, France

Abstract

POD curves are usually described using a parametric model which relates the probability to detect a specific defect to one of its geometrical characteristics, usually its size. The parameters of the model are estimated following a statistical procedure applied to a set of inspection results which are obtained thru dedicated experimental campaigns. Statistical significance requires that 60 to 80 mock-ups containing realistic flaws are fabricated and inspected by several inspectors. This costly and time consuming process must be done for each NDT procedure for which a measure of NDT reliability is required. Consequently, cost and time reduction of POD trials is currently a major issue. One way to achieve cost reduction is to replace some of the required experimental data with numerical simulation results. This idea follows the concept of Model Assisted POD (MAPOD). POD curves are no longer estimated from a fully empirical dataset but rather from a mix of experimental and simulated data. Simulations are performed using physics-based models, whose predictions are validated for the considered application case. In order to make the approach suitable for industrial needs, it is required that uncertainties introduced in the process thru the merging of simulation and experimental data are assessed.

In this presentation, a statistical method based on Bayesian updating is proposed, which mixes numerical simulations and information brought by the measurements. Traditionally, POD curves are assessed using Maximum Likelihood Estimation methods using either hit/miss or signal response data. This article only deals with hit/miss data. Following Berens article, the POD is modelled by a log-logistics function. Hit/miss data is treated as a Bernoulli's variable and Bayesian updating is performed on the POD model to assess the posterior distributions of the POD parameters assuming non-informative prior distributions on them. Finally, Monte Carlo simulations are run to assess the confidence band POD. A practical implementation of the approach to a high frequency eddy current inspection for fatigue cracks is presented.

References

[1] F. Jenson, N. Dominguez, P. Willaume and T. Yalamas, “A Bayesian approach for the determination of POD curves from empirical data merged with simulation results”, *AIP Conf. Proc.* 1511, 1741 (2013), http://dx.doi.org/10.1063/1.4789251

Merging of Experimental and Simulated Data Sets with a Bayesian Technique in the Context of POD Curves Determination

Speaker:

B. Chapuis[(1)]

Authors:

F. Jenson[(1)], N. Dominguez[(1)], P. Willaume[(2)], T. Yalamas[(2)]

(1) CEA, LIST, France

(2) Phimeca, France

October, 7th 2013

Background

- Inspection reliability is one of the key issues in ensuring safety of critical structural components
- Increasing use of probabilistic approaches based on statistical criteria such as POD curves and PFAs
- Are currently obtained thru expensive and time consuming experimental campaigns

- Approaches were proposed to replace some of the experimental data with simulation results

See:
MAPOD working group at www.cnde.iastate.edu/mapod
The PICASSO european project at www.picasso-ndt.eu

Objectives

Objectives of the study

- Define a Bayesian approach to POD curve determination (current approach relies on “classical” frequentist statistics)
- Determine POD curves from empirical data merged with numerical simulation data

Definitions	Empirical ~ experimental
	Data ~ inspection results

Outline

A Bayesian framework for the determination of POD curves

- Estimation of POD curves using a functionnal form (Berens approach)
- Focus on the hit/data format case
- Description of the Bayesian updating scheme

Bayesian approach to data merging

- First step: Bayesian update with simulation data
- Application case: HFET of fatigue cracks in Titanium
- Simulation supported POD curve determination (previous results)
- Second step: Bayesian update with empirical data

Conclusions and future work

Description of the Bayesian framework
Estimation of the POD curve using a functional form

- POD curves relate the detectability of a flaw to its size (or to another geometrical characteristic)
- Usual approach (MIL-HDBK 1823) consist in:
  - assuming a functional form for the POD curve
  - estimating the parameters of the function from the inspection results
  - estimating the associated confidence bound

Berens
- Hit/miss data format: *Log-odds function*
- Signal response data format: *Cumulative log-normal distribution function*

Cheng
- Estimated from a finite size data sample → sampling uncertainty → confidence bound

Description of the Bayesian framework
Estimation of the POD curve for hit/miss data

Hit – miss data: log odds model $$POD(a) = \frac{\exp(\alpha + \beta \cdot \ln a)}{1 + \exp(\alpha + \beta \cdot \ln a)}$$

Which can also be expressed as: $$\ln\left[\frac{POD(a)}{1 - POD(a)}\right] = \alpha + \beta \cdot \ln(a)$$

Estimation of α and β based on the principle of maximizing the likelihood function of the parameters given an observed sample of measurements (a_i, $a_{1/0}$)

The likelihood function of the parameters given the data is: $$L(\alpha, \beta) = \prod_{i=1}^{n} \prod_{j=1}^{k} p_i^{Z_{ij}} (1 - p_i)^{1 - Z_{ij}}$$

It indicates how likely chosen values of the parameters are, in light of the observed outcome

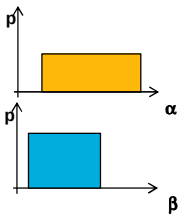
- n cracks
- k inspections
- Z_{ij}=0 or 1 for inspection j of crack i
- p_i is the probability value to detect crack i. The value is given by the log odds model:

$$p_i = POD(a_i)$$

Description of the Bayesian framework
Bayesian updating scheme

Assumptions

- Prior on α: **to be defined**
- Prior on β: **to be defined**



Non informative

Bayesian updating

Posterior is proportional to prior times likelihood

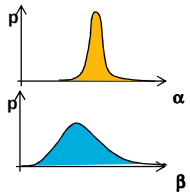
$P(A|B) \propto P(A) \times P(B|A)$

Likelihood function:
The variable follows a Bernouilli law

Data:
(crack size/ outcome of the inspection)
0 or 1 for « Hit-miss » values

Expected results

- Posterior law (sample) for α
- Posterior law (sample) for β



The objective is to update the knowledge on the parameters α and β

In order to solve the problem, the Monte-Carlo Marko Chains technique is used
The objective is to produce a sample which is representative of the posterior law

Outline

A Bayesian framework for the determination of POD curves

- Estimation of POD curves using a functionnal form (Berens approach)
- Focus on the hit/data format case
- Description of the Bayesian updating scheme

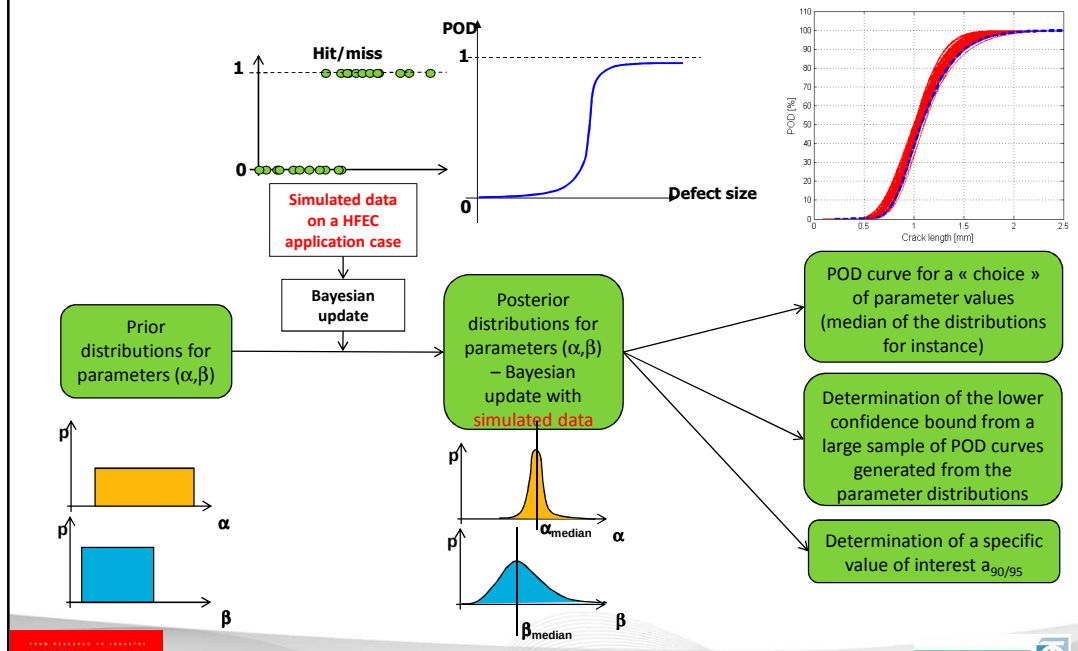
Bayesian approach to data merging

- First step: Bayesian update with simulation data
- Application case: HFET of fatigue cracks in Titanium
- Simulation supported POD curve determination (previous results)
- Second step: Bayesian update with empirical data

Conclusions and future work

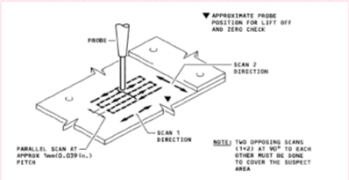
Bayesian approach to data merging

First step: BU with simulation data



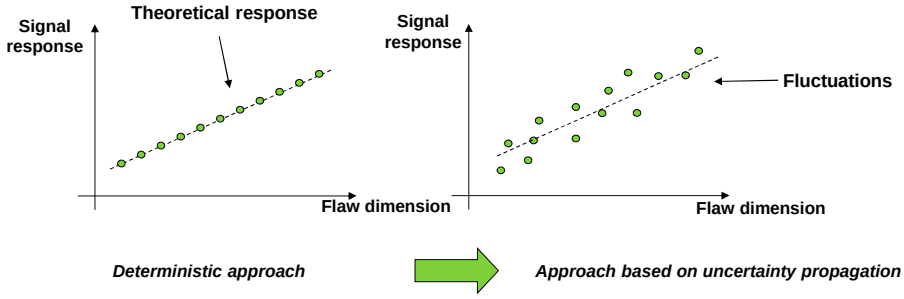
Bayesian approach to data merging

Application case: HFET of fatigue cracks in Titanium

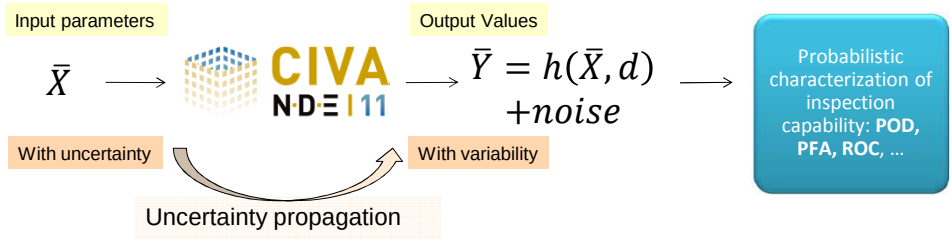
Part	NDT
Material: Titanium (TA6V) **Geometry:** Flat areas **Defects:** Fatigue cracks	**Configuration:** High Frequency Eddy Currents Testing (HFET)  **Probe:** Pencil probe (2MHz)  **Conditions:** In-service (manual)

Aim of the approach: simulate a set of "realistic values" for the NDT system response

- « realistic values»: values that take into account uncertainty and noise sources and thus reflect fluctuations around the theoretical responses



POD's in CIVA Software: a sampling approach based on existing physical models

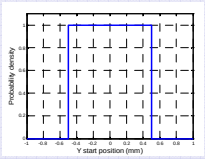
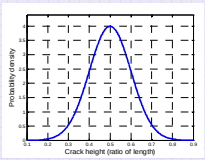
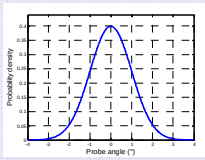
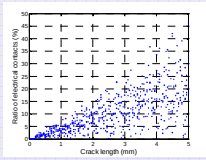


1. Definition of the inspection setup
2. Description of uncertainties on a set of input parameters
3. Propagation of uncertainty thru existing models (Monte Carlo)
4. Extraction of the output (the variable of interest)
5. Evaluation of probabilistic criteria such as the POD, PFA

Bayesian approach to data merging
Design of Numerical Experiments HFET on Titanium

Characteristic variable: crack length (mm)

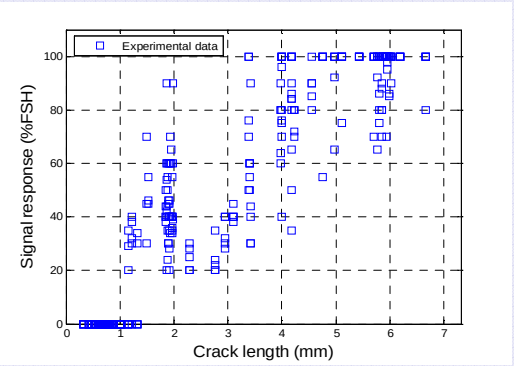
Uncertain parameters:

Start scan position	Crack height (mm)	Angle of the probe (°)	Electrical contacts in the crack
Corresponds to the position of the probe for picking the maximum amplitude signal	Fatigue cracking is subject to many uncertainties	Translated into an additional lift-off using geometrical rule	Real cracks exhibit electrical contacts
			
Uniform in [-0.5;05] (scan increment=1mm)	Gaussian 0.5*length +N(0,1)*0.12*length	Gaussian(0°;1°)	

Sampling strategy: Monte Carlo

Bayesian approach to data merging
Simulated and Experimental data

- **Experimental database**
  - 69 cracks from 0.33 mm to 6.66 mm
  - 5 operators



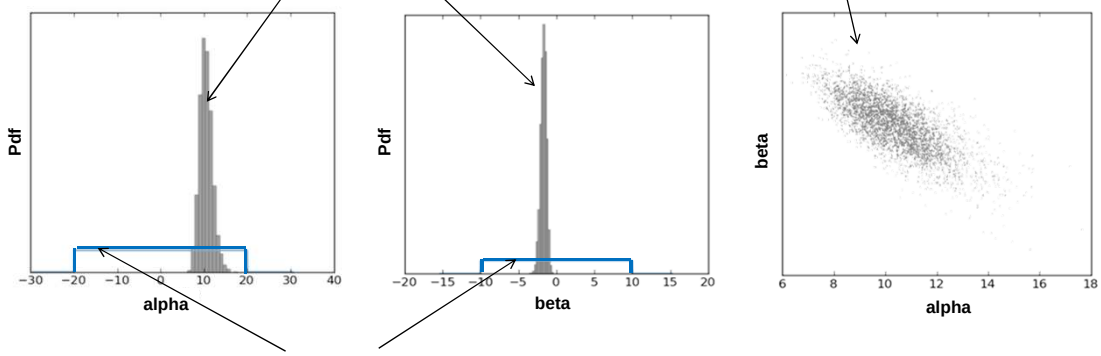
- **Simulations**
  - 100 crack lengths from 0.25 mm to 5.0 mm
  - 6 samples per crack length

Bayesian approach to data merging

Bayesian update on simulated data

Posterior distributions for alpha and beta

Sample generated thru MCMC

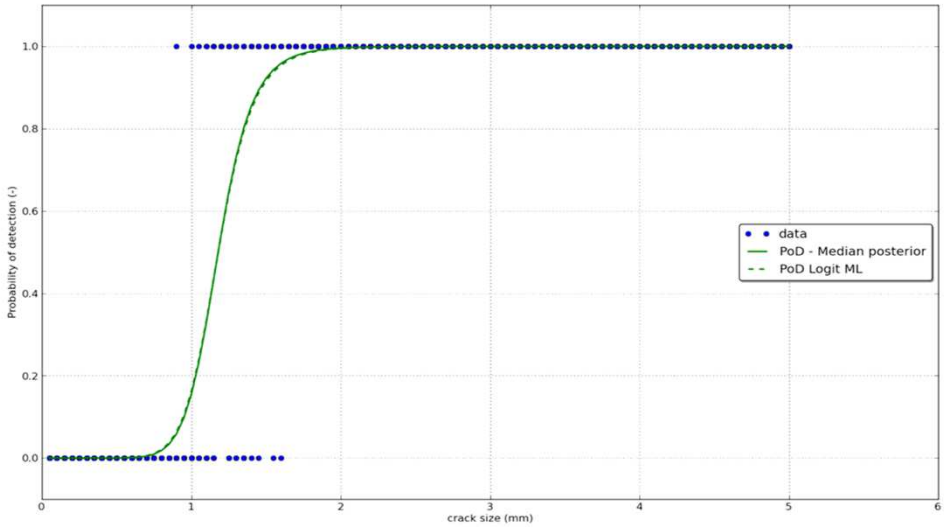


Prior distributions for alpha and beta (non informative)

- **Posterior distributions are centered and close to Gaussian distributions (slightly skewed)**
- **Sample of parameters values exhibit a statistical dependency**

Bayesian approach to data merging

POD Maximum Likelihood estimate (Berens) vs Bayesian Update



POD curves are almost identical

Bayesian approach to data merging

Computation of the POD curve with confidence

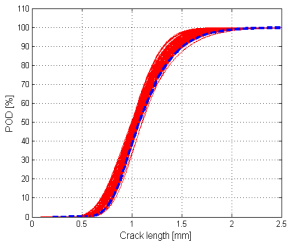
Definition:

The 95% lower confidence POD curve is the curve which leaves on the left 95% of the « POD curves » of the beam.

In practice the POD curve with confidence can be built pointwize, looping on POD values:

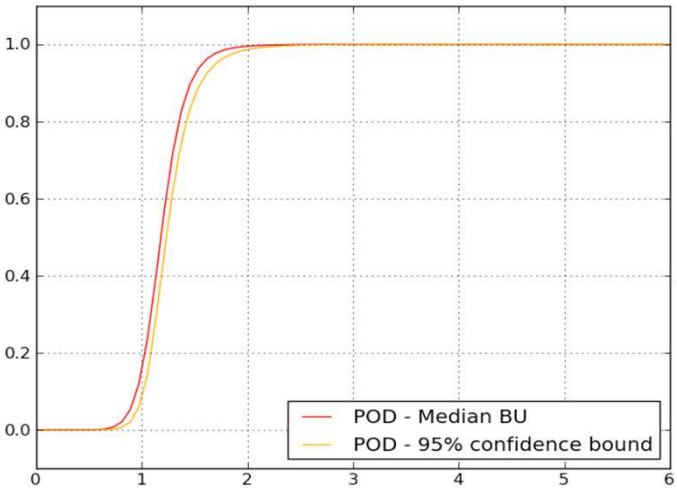
For p in [0.01;0.99]
  Find the POD curve leaving 95 % of the curves on the left at ordinate p
  Pick-up the corresponding flaw size a_p
End

The POD curve with confidence is the set of points (a_p, p).



Bayesian approach to data merging

Lower confidence band at 95%



Confidence bound close to the « median » POD curve due to the large size of the simulated data set (600 simulated inspections)

Outline

A Bayesian framework for the determination of POD curves

- Estimation of POD curves using a functionnal form (Berens approach)
- Focus on the hit/data format case
- Descirption of the Bayesian updating scheme

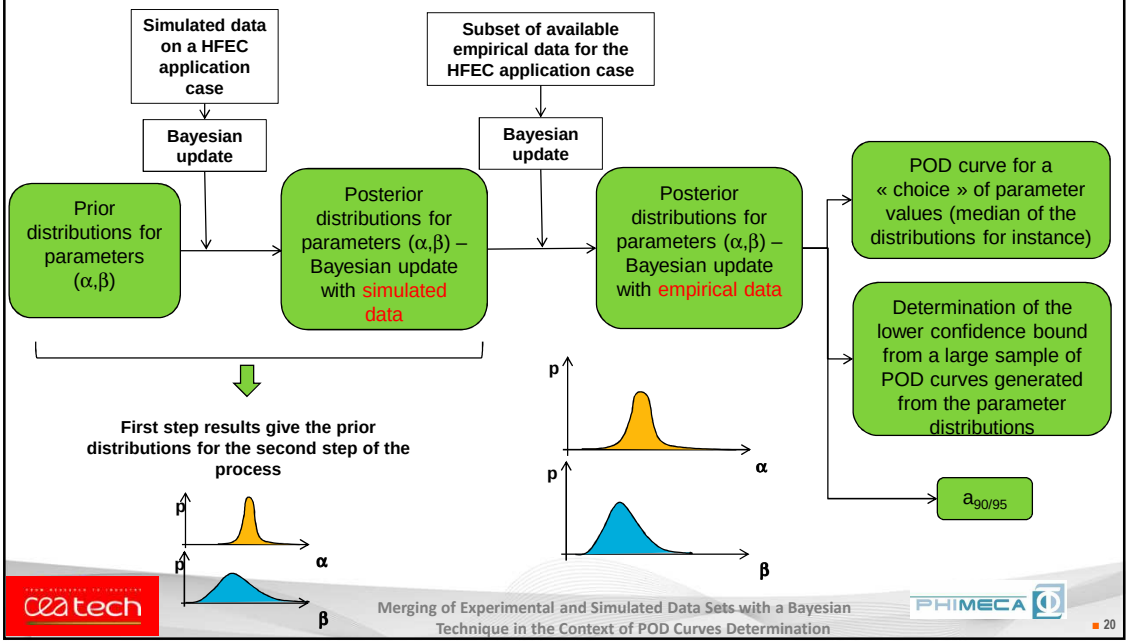
Bayesian approach to data merging

- First step: Bayesian update with simulation data
- Application case: HFET of fatigue cracks in Titanium
- Simulation supported POD curve determination (previous results)
- Second step: Bayesian update with empirical data

Conclusions and future work

Bayesian approach to data merging

Second step: Bayesian update using empirical inspection results

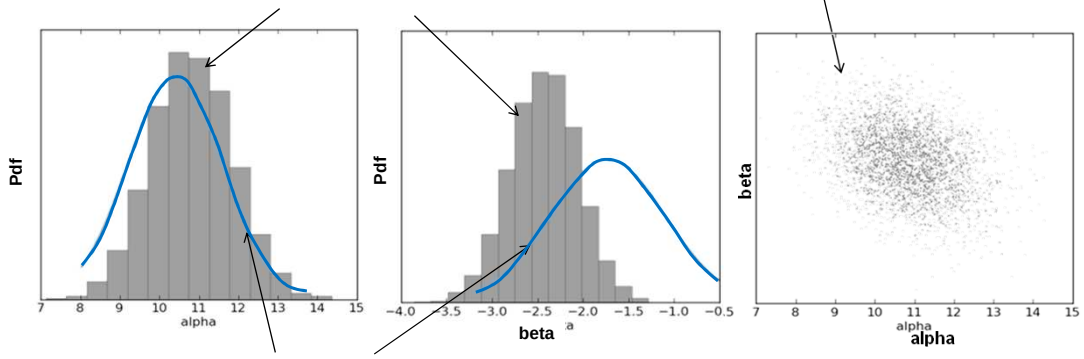


Bayesian approach to data merging

Bayesian update on empirical data

Posterior distributions for alpha and beta

Sample generated thru MCMC

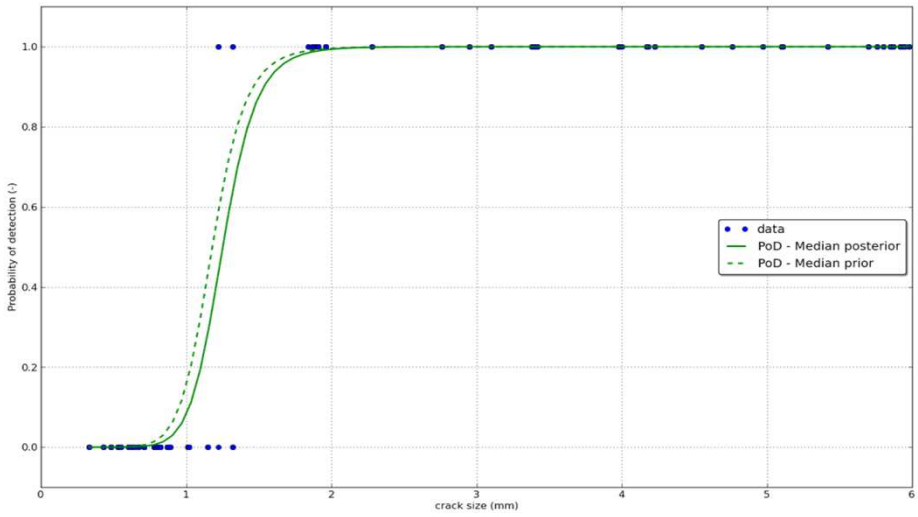


Prior distributions for alpha and beta (obtained thru BU on simulated data)

- **Mean values of the posterior distributions slightly different**
- **Variance of the posterior distributions are reduced**

Bayesian approach to data merging

POD BU on simulated data vs. BU on empirical data



POD curves are very close
BU on experimental data moves the curve to the right

Outline

A Bayesian framework for the determination of POD curves

- Estimation of POD curves using a functionnal form (Berens approach)
- Focus on the hit/data format case
- Descirption of the Bayesian updating scheme

Bayesian approach to data merging

- First step: Bayesian update with simulation data
- Application case: HFET of fatigue cracks in Titanium
- Simulation supported POD curve determination (previous results)
- Second step: Bayesian update with empirical data

Conclusions and future work

Conclusions

Bayesian approach to the determination of POD curves

- Bayesian approaches are “richer” than MLE: we obtain distributions on the input parameters, not just a single value
- For a given choice of the parameters (e.g. the median values), the resulting POD curve is very close to the Berens (MLE) one

Application of the method to HFET of titanium plate

- Numerical model results are close to experience: the second update done with experimental data doesn’t modified significantly the prior that was obtained with the simulation results
- The parameter variances are reduced

Perspectives

- Demonstrate that the method allows minimization of the required empirical data set

- Apply Bayesian techniques to other problems:
  Definition of the input parameters uncertainties in the simulation supported approach to POD curve determination

- Application to the signal response data format

- Show that less empirical data are required to achieve a certain confidence level than what is demanded by recommendation documents (MIL-HDBK 1823a)

Thank you for your attention!

Description of the Bayesian framework

How to define prior distributions for the parameters of the POD model?

Interpretation of the parameters α and β

α and β are not easily interpretable in physical terms ⇨ How to evaluate prior knowledge?

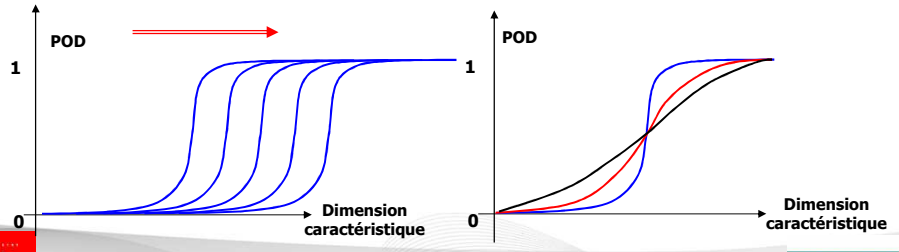
An equivalent form of the logit model is: $$POD(a) = \left\{1+\exp\left(-\frac{\pi}{\sqrt{3}}\left(\frac{\ln a-\mu}{\sigma}\right)\right)\right\}^{-1}$$

With: $\mu=-\alpha/\beta$, $\sigma=\pi/(\beta.\mathrm{sqrt}(3))$

In this form:

$\mu=\ln(a_{0.5})$ where $a_{0.5}$ is the flaw size which is detected with a probability of 50%

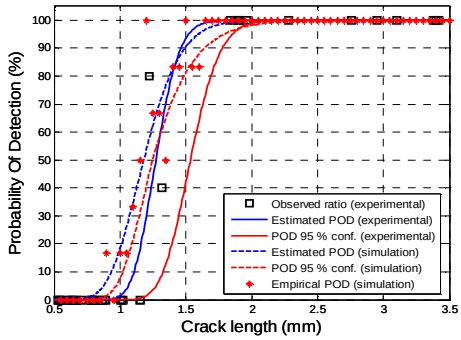
σ is the steepness of the POD(a) function. The smaller the value of σ, the steeper the POD(a) function



Bayesian approach to data merging

Simulated vs. Experimental POD curves

- 600 simulated data
- 4 uncertain parameters (start position, crack height, probe angle, elec. contacts)



POD curves very similar

- Slope corrected but still the effect of the number of data
- Confidence band smaller for simulation dataset because more data than in the experimental dataset

$$\begin{cases} a_{90}^{\exp} = 1.5mm \\ a_{90/95}^{\exp} = 1.8mm \end{cases} \qquad \begin{cases} a_{90}^{simu,EC} = 1.5mm \\ a_{90/95}^{simu,EC} = 1.7mm \end{cases}$$

Values of interest very similar