MOX–Report No. 95 – dicembre 2006

Reliability of Computational Science

I. Babuška, F. Nobile, R. Tempone

MOX, Dipartimento di Matematica "F. Brioschi"
Politecnico di Milano, Via Bonardi 9 - 20133 Milano (Italy)

mox@mate.polimi.it http://mox.polimi.it

Reliability of Computational Science

I. Babuška[1*] F. Nobile[2†]

R. Tempone[3‡]

[1]ICES, University of Texas at Austin, Texas, USA
[2]MOX, Dipartimento di Matematica, Politecnico di Milano, Italy
[3]SCS and Deparment of Mathematics, Florida State University, USA

December 12, 2006

Abstract

Today's computers allow us to simulate large, complex physical problems. Many times the mathematical models describing such problems are based on a relatively small amount of available information such as experimental measurements. The question arises whether the computed data could be used as the basis for decision in critical engineering, economic, medicine applications. The representative list of engineering accidents occurred in the past years and their reasons illustrates the question. The paper describes a general framework for Verification and Validation which deals with this question. The framework is then applied to an illustrative engineering problem, in which the basis for decision is a specific quantity of interest, namely the probability that the quantity does not exceed a given value. The V&V framework is applied and explained in detail. The result of the analysis is the computation of the failure probability as well as a quantification of the confidence in the computation, depending on the amount of available experimental data.

1 Introduction

Computational Science is a discipline concerned with the use of computers for the prediction of physical phenomena. These predictions are used as the basis for critical decisions in engineering and in other fields such as environment, heath,

*Partially supported by Sandia National Lab. grant N. 268687

†Partially supported by J.T. Oden Visiting Faculty Fellowship Research Program, ICES, University of Texas at Austin and M.U.R.S.T. Cofin 2005 "Numerical Modeling for Scientific Computing and Advanced Applications"

‡Partially supported by Sandia National Lab. grant N. 268687 and J.T. Oden Visiting Faculty Fellowship Research Program, ICES, University of Texas at Austin

management, etc. Rapid development of computer hardware allows us to make predictions of more and more complex phenomena.

The major problem arises: How reliable are these predictions? Could they be the basis for decisions, often very crucial and with large implications? The reliability problem has many aspects: mathematical, numerical, computational, experimental, and philosophical. The present paper addresses some of these aspects in a general way and particularizes them to specific applications.

We will start with few examples of specific engineering accidents and their reasons. We divide them into four categories: A) the modeling, B) the numerical treatment, C) computer science problems and D) human errors.

A. Modeling problem

- *The Tacoma Narrows Bridge.* The suspension bridge across Puget-Sound (Washington State) collapsed November 7, 1940. Reason: the model did not properly describe the aero-dynamical forces and the effects of the Von Karman vortices. In addition, the behavior of the cables was not correctly modeled.
- *The Hartford Civic Center roof (Connecticut).* It collapsed January 18, 1978. Reason: linear model and models of the joints were inadequate.
- *The Columbia Shuttle Accident June 2003.* It was caused by a piece of foam broken off the fuel tank. After it was observed, the potential of the damage was judged, upon computations, as non-serious. Reason: the model used did not take properly into consideration the size of the foam debris.

B. Numerical treatment problem

- *The Sleipner accident.* The gravity base structure of Sleipner, a off-shore platform made of reinforced concrete, sank during ballast test operation in Gandsfjorden, Norway, August 23^{rd}, 1991. Reason Finite element analysis gave a 47% underestimation of the shear forces in the critical part of the base structure.

C. Computer science problem

- *Failure of the ARIANE 5 rocket, June 1996.* Reason: problem of computer science, implementation of the round offs.

D. Human error

- *Mars Climate Orbiter.* The Orbiter was lost September 23, 1999, in the Mars Atmosphere. Reason: unintended mixture of English and metric units.

There were many more accidents and mishaps. We wanted to mention only a few, as representatives of the main four categories. Today, in most cases, the bottleneck of reliability is the proper modeling of the physical system, giving rise to problems in the category A. This paper elaborates mostly on this category but also briefly on the category B. The categories C and D will not be addressed here. The category B problems relate directly to the mathematics, while category A is broader.

The question of the reliability of predictions based on mathematical models has received much attention in many fields of applications for a long time. Nevertheless now, thanks to the availability of large computers, the problem of the reliability is becoming more and more important. There is a vast literature available but the field is still developing. The reliability problem is directly related to the Verification and Validation (V&V) field.

This paper has essentially two parts. In Part I, the section 2 is addressing a sample of the literature on the subject of V&V while Sections 3-8 formulate the basic notions and main ideas and their philosophical underpinning. Part II (Sections 9 on) applies and implements the main ideas of Part I on a specific problem of the reliability of a frame. We address the reliability of the computational prediction which is based on a small number of available experimental data used to determine the input parameters of the numerically solved mathematical problem.

The goal of the paper is to show some basic ideas of the V&V field and an application of these ideas to an academic engineering problem when only limited information is available. The ideas and methodology presented in the paper are general and applicable to much more complex problems and not only those of an engineering type. The paper is in some sense an overview paper referring to [BNT06] for various details. We hope that the richness of open problems of mathematical character in the V&V field will be apparent.

Part I
General V&V framework

2 Selected literature on V&V

Here we will mention a small sample of relevant papers on V&V.

- [AIA98] is the report of AIAA which was one of the first addressing V&V. It is a very influential report and is very often cited.
- [OT02] is a survey paper with a large list of literature.
- [Roa98] is a good, easily readable book on the subject addressing in details main ideas of Verification (mostly based on Richardson extrapolation) and Validation.
- [CF99] is a book surveying the probabilistic techniques related to the validation. The book contains a large list of references.
- [Pos04] is a very interesting and essential paper. It formulates and discusses three basic challenges of the Computational Science: A) performance, B) programming and verification, C) modeling and validation. The paper arrives at the conclusion that the modeling (validation) problem is the major challenge and the bottleneck of the success or crisis of computational science.
- [BS01] is a voluminous book addressing the reliability of the finite element method, a-priori error estimates, convergence, pollution, superconvergence, a-posteriori estimates. Many numerical computational illustrations are presented.
- [HCB04] is a book focusing on the effects of uncertainty in the input data on the solution. It is a mathematical book addressing the worst scenario approach. The book has a large introductory chapter about several approaches to treat uncertainty.
- [BO04] presents the basic notions of Verification and Validation and their implications.
- [BO05] is more or less a survey paper on V&V with specific characteristic examples: a) The linear elasticity problem when the material coefficients are given by the fuzzy sets. Mathematics of it is addressed in [BNT05]. b) The problem of the constitutive law for cyclic plasticity discussed in light of extensive experiments addressed in [BJLS93]. c) Stochastic formulation and its application, which are elaborated in detail in [BTZ04, BTZ05].

- [BNT05] addresses the problem when the only information on the coefficients of the partial differential equation is their range. Given the quantity of interest the problem is to obtain its range, a-posteriori error estimation and the coefficients leading to the bounds of the range.

- [BJLS93] addresses the reliability of the constitutive law for cyclic plasticity in the light of one dimensional experiments. Large number of experiments were performed and statistically analyzed. Among others it was shown that the classical constitutive law used in engineering is very unreliable.

- [BTZ04] addresses theoretical foundations and convergence of the finite element method for stochastic partial differential equations.

- [BTZ05] addresses the adaptive FEM for solving stochastic PDEs. Introductory chapters give a survey of various aspects of treatments of problems with uncertainties.

- [BLT03] analyzes the question of solving simple bar problems based on the use of various number of the experimental data of the Young's modulus of elasticity. The statistics of the experimental data is used in the Karhunen-Loève formulation, which is at the basis of the probabilistic description of the problem.

- [Bab61] is a very old paper addressing the stochastic solution of the Laplace equation with stochastic right hand side. This methodology was used for the computational analysis of certain aspects (concrete freezing) related to the building of the dam Orlik in early fifties in Czechoslovakia.

- [BNT06] is the basis of the second part of the present paper. It is related to the Validation Challenge problem, Sandia National Laboratory, May 2006.

- [NF67] is an old paper, mostly addressing economic problems where validation issues have been of interest for a long time. It discusses philosophical aspects and has influenced many simulation textbooks.

- [KOG98] is very general with examples mostly related to the economy. It addresses various philosophical foundations of the validation. The approach of the validation of the frame problem addressed in this paper is philosophically in the direction of Methodological Falsification in the sense of Popper and Lakatos.

3 Basic notions

The purpose of computation is to provide the quantitative data of interest (sometimes called quantities of interest) on which a decision is made.

These quantities are predictions of certain phenomena relevant for the decision. Decision making techniques, as for example the utility theory, are not discussed in the present paper.

Sometimes the computation is made only for understanding certain phenomena and only qualitative characterization is of interest. We will not elaborate on it in this paper

The scheme of the Computational Science approach is shown in the Fig. 1. "Mathematical Model" is a synonym for "Mathematical Problem". Its relation

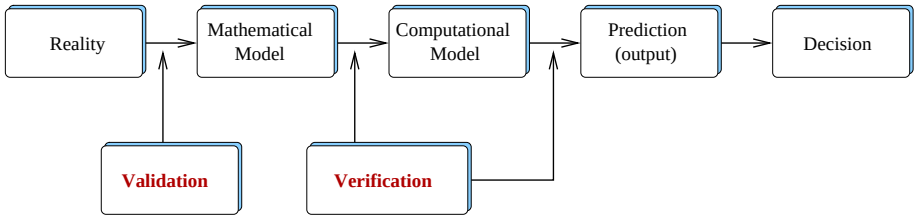


Figure 1: Scheme of the Computational Science

to the reality is the Validation problem.

The mathematical model only transforms the available information into the prediction of the quantity of interest.

Hence the reliability of the prediction depends on the quality of the available information. The mathematical problem is then solved by a numerical approach, which creates a computational model. The relation between the solution of the mathematical and the computational models is the subject of Verification. Accidents of the category A mentioned in Section 1 are related to the Validation process. The accidents of the categories B, C, D are related to the Verification process.

4 Mathematical Problem

The scheme of the mathematical problem of interest is shown in Fig. 2. The structure of the problem could be for example an elliptic PDE and the Newton boundary condition. The input data are the coefficients of the PDE, the functional form of the right hand side and boundary conditions as well as the physical domain. The output could be the value of a functional, for example the value of the solution in a particular point. The problem, the admissible sets (spaces) of the solution and the input data have to be well defined. The output

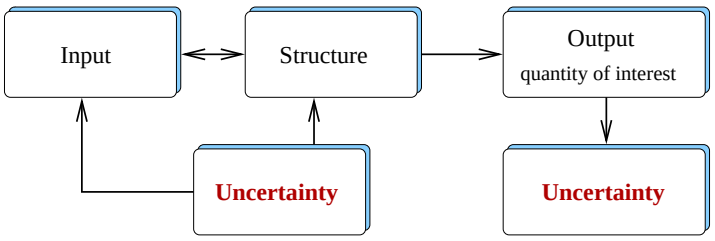


Figure 2: Scheme of the mathematical problem

has to be properly defined so that it has a proper sense with respect to the space of admissible solutions. An essential part of the mathematical problem is the definition of the quantities of interest or other goals for computations. The mathematical problem has to have reasonable properties, for example the existence of the solution, its continuous dependence on input data etc.

The mathematical model defines a *general* problem. When endowed with input data (for computational analysis) then the problem become *specific* and is used for the prediction. The specific (prediction) problem reflects the available information and its character. It can be deterministic, stochastic, worst scenario type etc.

5 Quantification of the uncertainties

We said that the mathematical problem creates the transformation of the available information into the desired ones. This information is obtained by experiments, experience, expert opinions and always contains uncertainties. They have to be quantitatively specified via probability fields, fuzzy sets, ranges (for the worst scenario approach) etc. This specification is usually not easy because not enough experimental data is available. The mathematical problem, i.e. the input, the structure and the output, reflects the character of the uncertainties and their quantitative description. This description is directly related to the goal of the computation (quantity of interest). Sensitivity analysis plays here an important role: it influences the decision on which uncertainties in the input data have to be retained and which ones can be neglected, instead.

The uncertainty can be aleatory or epistemic. The aleatory uncertainty is related to the physical uncertainty and cannot be decreased or avoided. The epistemic uncertainty (called sometimes the ignorance) can be in principle avoided by better experimental technology, better understanding etc. Nevertheless, in practice, it cannot be totally eliminated either. The quantitative description of the aleatory and the epistemic uncertainties is usually different and this is also reflected in the formulation of the mathematical problem.

6 Calibration

To identify all or part on the input data in the specific mathematical problem of interest, we select suitable calibration problems. These have to be related to the goal of the analysis (i.e. the prediction) and could be both experimentally and numerically analyzed via a possibly simpler mathematical model.

The input data are selected so that a good agreement between the results based on the model and the experiments is obtained in the specific calibration problems. The determination of these input data is called the *calibration*. The calibration problems can have a deterministic or stochastic nature and have to be relevant for the problem of interest. Let us consider for example the three dimensional problem of plasticity and cyclic loading. The major part of the problem is the specification of the constitutive law. Then the (three dimensional) constitutive law is selected and calibrated so that results are in good agreement with one dimensional experiments. (This of course does not mean that the law will be good in three dimensions. This is the subject of validation).

The calibration experiments are relatively cheap. The calibration is always based on experiments that are different from the prediction problem, because the prediction problem cannot be experimentally analyzed. (It seldom occurs that, after the computational analysis has been performed, experimental measurements become available for the prediction problem. In this case we speak about *post-audit*. The post audit analysis is typically done when an accident occurs). The specific mathematical problem with the input data based on the calibration is then addressed in the validation phase.

7 Validation

> *Validation is a process determining if the mathematical model describes sufficiently well the reality with respect to the decision which has to be made.*

The validation process usually is related to the validation pyramid of experiments with increasing complexity approaching the prediction. The cost of the validation experiments is increasing with their complexity. Hence the number of the available experiments decreases with their complexity. In the Fig. 3 we show an idealized validation pyramid, which is related to an aircraft structural design. Other validation tests outside of the pyramid are often added. We are in dept to Mr. Stéphane Guinard of European Aeronautic Defense and Space (EADS) Corporate Research for the permission to publish this figure. On the left hand side of the pyramid are the experiments with the increasing complexity. On the right hand side are the computational models. At the lowest level of the pyramid are the simple calibration experiments. On the highest level are very complex experiments and their computational analysis. Some of them are called accreditation (certification) experiments and serve as the basis for the

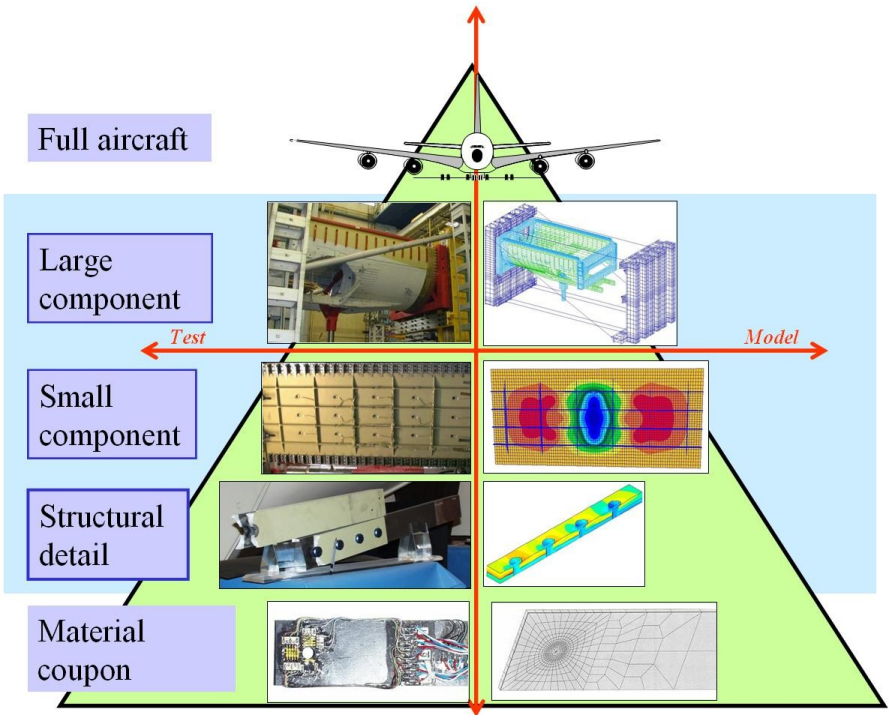


Figure 3: An idealized validation pyramid related to an aircraft structural design.

demonstration of compliance with regulatory requirements. Sometimes, tests on the lower level might be accreditation tests, too.

The comparison between the experimental (validation) data and the computed data is based on a specific metric (i.e. how the difference is measured) and the rejection criterion, which is a quantitative measure of the difference. The metric and the criterion have to be directly related to the prediction and the decision based on it. If the criterion is larger then the given tolerance, which is related to some threshold conditions, the model will be rejected. If the model at any level of the pyramid is rejected then the model has to be changed and to pass all the lower level tests and possibly more experiments would be needed. If the model is not rejected at a certain level of the validation pyramid, then the higher level is performed.

The used tolerance is not arbitrary. It relates to the required accuracy of the prediction. If the required accuracy is low then the tolerance could be large so that even a very crude model will not be rejected. If the desired accuracy is high then the tolerance has to be small and many models could be rejected. The tolerance has to be chosen reasonably, otherwise any practical model could possibly be rejected. If more than one model are calibrated and validated then the best model could be possibly chosen and the tolerance adjusted so that the model will not be rejected. This can be done only if the adjustment is

admissible for the decision based on the prediction. If the model is rejected then a new model has to be created.

The design of the validation pyramid is crucial. It is essentially an optimization problem: find the pyramid in the financial budget so that the reliability of the prediction is maximal.

To illustrate how the accreditation tests may lead to the rejection of the model, we mention the Airbus A380 test. The accreditation wing test failed on February 14, 2006. (Flight International 16/02/2006). EASA (European Aviation Safety Agency) specifies that the wing in the static test has to endure a load which is 150% of the limit load (worst scenario metric) for $3sec$. The wing broke at the point between the inboard and outboard engine at the 147% of the limit load. Some adjustment of the wing design is expected.

Airbus Executive Vice President of Engineering A. Garcia said at the press conference: *This is within 3% of the 1.5 target which shows the accuracy of the finite element analysis.*

In this connection we also cite the comment of J. Kirby (see Flight International): *No computer code is 100% accurate. Tests are required to verify code predictions. Hopefully the test data show that the code predictions are conservative. But premature failure is not necessarily a disaster. I recall that on one test program we had a premature failure (also a structural test). We found with the aid of the test data that we had not modeled one aspect of the design correctly. Correcting the computer model showed that the premature failure was predictable. That gave us the confidence to modify the design so that we met the ultimate load.* Although the terminology in this citation is not completely the one we are using in our present paper it shows nicely the rejection based on the validation test and the change of the mathematical model.

By the validation test we can only reject a model based on a particular tolerance. It does not necessarily mean that the prediction model not rejected is an accurate description of the reality in the range of the given tolerance.

Validation is an induction process. If a model is not rejected on an increasing number of validation experiments our confidence in the prediction grows. This process is very closely related to the philosophy of Popper and Lakatos, namely their Methodological Falsification (see [KOG98]).

Let us mention that the accreditation test is part of the legal process and is not necessarily a characterization of an engineering accuracy. For example, obtaining a value of 1.47 instead of 1.5 in the Airbus A380 test, could be sufficient from the engineering point of view, to access the accuracy of the computational model in the prediction. Nevertheless, it is a clear failure from the regulatory standpoint. The objective, taking into account the regulatory constraints is to achieve a larger value than 1.5.

8 Verification

Verification is a process of determining if the computational model and the implementation lead to the prediction with sufficient accuracy i.e. the difference between the exact and computed prediction is sufficiently small.

Verification consists of: a) mathematical part, i.e. the analysis of the numerical method, convergence, a-posteriori error estimation with respect to the desired output, (the quantities of interest). It is a purely mathematical process; b) the analysis of the correctness of the code. Here the manufactured solution technique is one important tool. Also finding errors in the input data and other subtle computer science aspects belong to the verification process.

Verification is also important in calibration and validation. We need to have a sufficiently accurate numerical solution of the problem which is compared with the experimental data, otherwise it would be impossible to analyze the reliability of the model. Verification of the computational model cannot be based on the comparison with the experimental data. Although in practice the computational model is compared with the experiments (see the citation of J. Kirby in the previous section), it is necessary to assume that the computational model was verified so that its error is negligible with respect to the measure of the difference between computed and experimental data.

Finally let us remark that verification applies also to the experimental work. Here we mean that the data which are measured are those which are needed and that their accuracy is sufficient.

Part II
An illustrative engineering problem

9 The frame prediction problem

In this and the following sections we apply the ideas of the previous sections to the analysis of the reliability of a frame. It is a simple academic problem which illustrates well the general ideas and addresses the general methodology of the solution approach. The problem is one of the Validation Challenge Workshop, Sandia National Laboratory, May 27-29, 2006, Albuquerque, NM. For detailed analysis we refer to [BNT06]. Fig. 4 shows the frame problem, whose dimensions are given in Table 1. The bar 4 is loaded by the uniform load of intensity $q = 6KN/m$.

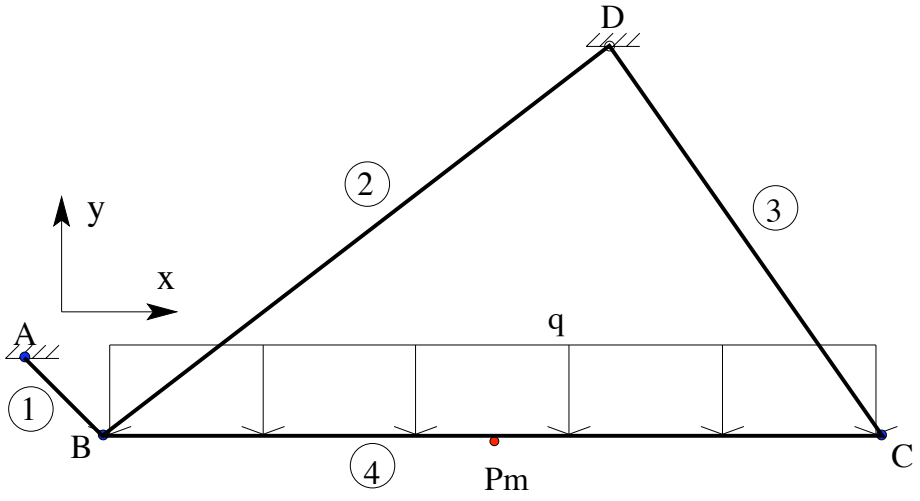


Figure 4: Prediction frame: structure and uniform load q under study. We are interested in the vertical displacement of point P_m.

Point	x(cm)	y(cm)	Bar #	$A(cm^2)$	$I(cm^4)$
A	0	20	1	16	
B	20	0	2	16	
C	220	0	3	16	
D	150	100	4	80	5333

Table 1: The geometrical data of the frame from Figure 4. Observe that only beam 4 is subject to bending.

The vertical displacement of the point P_m is the phenomenon of interest.

It is the basis of the decision on the frame reliability. The joints (hinges) are assumed to be perfect and their support in the points A and D is rigid. In reality the hinges are not perfect and the support is not completely rigid. Nevertheless, from a careful analysis of the design details and a simple sensitivity analysis, it was concluded that this idealization has no influence on the decision based on the displacement in the point P_m. The geometrical data are assumed to be completely accurate and the load q is given by the regulation. It was also concluded that the use of the Kirchhoff bending theory for the bar 4 is acceptable for the decision. The material property, specifically the Young's modulus of elasticity significantly influences the displacement and hence the reliability of the analysis of the frame. The mathematical model is linear. Input data and quantity of interest are given in Fig. 4 and Table 1. The input are the position of the hinges, the point P_m, the cross-sections and modulus of elasticity of the bars as well as the load. The structure of the problem is derived from the classical linear structural mechanics. The quantity of interest is the probability that the displacement $w(P_m)$ in the point P_m will not exceed $3mm$. The goal of the analysis is to give this probability and describe the confidence in the computed data.

10 Quantification of the uncertainty for the frame problem

It was concluded that the only uncertainty influencing the decision is the material property characterized by the Young's modulus of elasticity E. It is described in a probabilistic way. It will be assumed that the compliance $\mathcal{C} = 1/E$ is a stochastic function described by a stationary random field that is completely characterized by the marginal distribution of $\mathcal{C}$ and two additional parameters L_c and $\alpha > 0$ (defined in Section 12). The bars in the frame are assumed to be independent in the probabilistic sense. The information about the modulus of elasticity and its probability distribution is obtained from the calibration experiments. Because the number of experiments is small there is still uncertainty in the constructed probability field. It is assumed that all experimental measurements are perfect. Since the input data are described in a probabilistic way with uncertainty, the quantity of interest, namely the displacement in P_m will be also described in probabilistic terms with uncertainty. Further it is assumed that all used algorithms are verified.

11 The validation pyramid

The calibration, validation and accreditation experiments described bellow create the validation pyramid analogous to the one shown in Fig. 3.

a) Calibration experiments. The calibration experiments are the basis for

the specifics of the input data so that the specific (prediction) model will be defined. The experiment is the material coupon of cross-section $A = 4.0cm^2$, length $L = 20cm$ which is loaded by the force $F = 1.2KN$. In the middle of the coupon the strain is measured by a strain gage and the elasticity modulus is computed. In addition the elongation of the bar, δL is measured. See Fig. 5. For disposition are three groups of measurements with different number of experiments (samples): $N_c = 5$, $N_c = 20$, $N_c = 30$. Because the calibration experiments are cheap, more samples can be measured than for validation. Table 2 gives the elongation and the strain for the coupon samples.

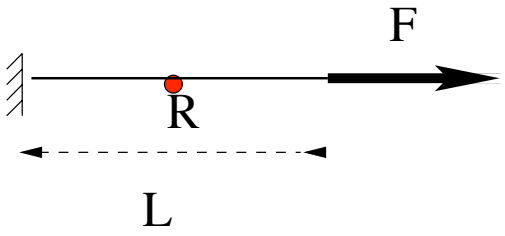


Figure 5: Scheme of the calibration experiments.

b) Validation. The validation is testing the specific model based on the calibration data. The model could be possibly rejected. The validation test is the bar of the length $L_v = 80cm$ with the cross-section $A = 4.0cm^2$, loaded by the force $F = 1.2KN$. The elongation δL_v is measured. To simulate higher costs of the validation tests we have for the three groups only a smaller number of samples, namely $N_v = 2$, $N_v = 4$ and $N_v = 10$. See Fig. 6. Table 3 gives the elongation measurements.

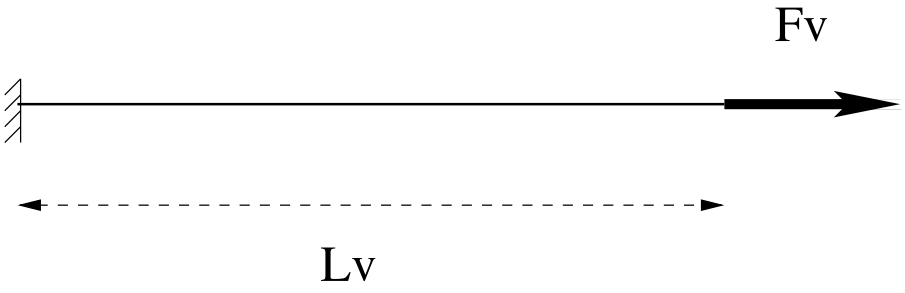


Figure 6: Scheme of the validation experiment.

c) Accreditation. The accreditation (certification) experiment is the most expensive test at the top of the pyramid, just below the prediction. The accreditation test is the frame shown in Fig. 7. It consists of four bars.

Sample #	δL(mm)	$E(L_c/2)$ (GPa)
1	5.15e-02	13.26
2	5.35e-02	10.86
3	5.24e-02	14.77
4	5.51e-02	10.94
5	5.14e-02	11.05
6	5.38e-02	11.06
7	4.97e-02	11.97
8	5.41e-02	11.66
9	4.95e-02	12.09
10	5.42e-02	11.30
11	5.47e-02	10.98
12	5.74e-02	11.92
13	5.36e-02	11.12
14	5.42e-02	12.00
15	5.34e-02	10.98
16	5.60e-02	10.71
17	5.06e-02	10.91
18	4.99e-02	11.89
19	5.22e-02	11.43
20	5.57e-02	10.87
21	5.28e-02	11.75
22	5.10e-02	13.47
23	5.48e-02	11.44
24	5.35e-02	12.44
25	4.92e-02	12.13
26	5.51e-02	11.38
27	5.27e-02	10.75
28	5.14e-02	11.92
29	5.61e-02	10.82
30	5.56e-02	11.04

Table 2: Measured elongation δL and the modulus of elasticity $E(L/2)$ in the calibration experiments

The bar 1 is bended by the concentrated force $F = 6KN$ located in the middle of the bar. The displacement is measured in the point under the load. The bars 2 and 4 are not attached at their crossing. The geometrical data of the frame are given in Table 4. Because of the high cost of the accreditation test we have for disposition only $N_a = 1$ sample in the group 1 and 2 and $N_a = 2$ accreditation frames in the group 3. Table 5 gives the displacement under the load.

Sample #	δL(mm)
1	2.01e-01
2	2.06e-01
3	2.01e-01
4	2.08e-01
5	2.04e-01
6	2.01e-01
7	2.06e-01
8	2.11e-01
9	1.98e-01
10	2.08e-01

Table 3: Elongation measurements for the validation test.

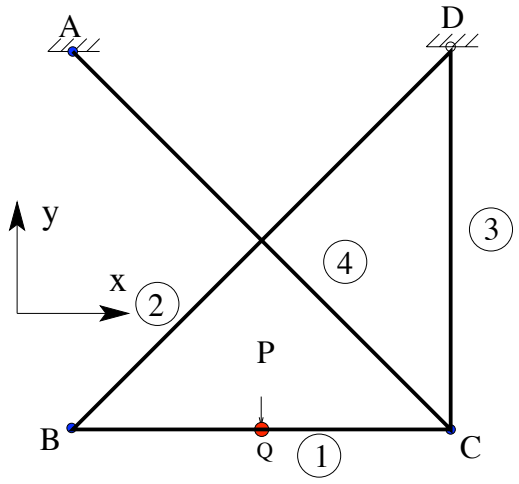


Figure 7: The accreditation test frame.

Point	x(cm)	y(cm)	Bar #	$A(cm^2)$	$I(cm^4)$
A	0	50	1	16	333.3
B	0	0	2	16	
C	50	0	3	16	
D	50	50	4	20	

Table 4: Geometrical data of the accreditation test.

12 The methodology of the approach

The goal of the analysis is to give the probability that the displacement at the point P_m of the frame will not exceed $3mm$. Also the confidence in the com-

Sample #	$w(P)$(mm)
1	-6.50e-01
2	-6.73e-01

Table 5: Measured displacements under the load for the accreditation tests. The first measurement is used for cases 1 and 2

puted probability has to be given. For disposition are data from the calibration, validation and accreditation experiments. To see how the results are influenced by the number of experiments, three sample sets of data, described in the previous section, are analyzed. The ideas and the methodology used in the paper are general and are not restricted to the academic frame problem.

First, we design the calibration experiments from which we will obtain the probability field of the compliance. We consider various marginal probability distributions, parametric and non parametric and fit the data. We use the bootstrapping approach [Efr82, ET93] to give variability bounds for the fitted parameters. After calibration is performed, we rank the marginal probability distribution models by the Kullback-Leibler discrepancy theory [KL51]. Due to the small amount of validation data, we adopted a Bayesian approach to include the new available information and produce an updated (possibly better) model. Then, the metric and criterion for rejection is based on the distance between the prediction of the quantity of interest (displacement of the point P_m of the frame) using the calibrated model and the Bayesian updated one. A similar Bayesian approach is used for the accreditation metric and criterion.

The calibrated model, if not rejected, is then used for the prediction. The confidence in the computed prediction is based on the computed distance between the calibration and validation, resp. accreditation data. We also include in the prediction information on the variability of the calibrated input data, estimated by bootstrapping, to account for the fact that only a small number of experiments is available.

We do not use simultaneously all the available data, i.e. calibration, validation and accreditation for the prediction. We only use the calibration data for the prediction, while the validation and accreditation data are used to characterize the confidence in the prediction. The main reason is that in a complicated problem as for example the airplane design (see Fig. 3) the simultaneous use of all the data from the validation pyramid is practically impossible. The metric we are using here is based on the frame prediction problem. In other words, the mismatch between the computed and experimental validation (resp. accreditation) data is "mapped" onto the prediction and the metric and rejection criterion are defined directly at the prediction level. For a more complex problem, this "map" can be very difficult to achieve and one could use a surrogate simplified model, instead.

We consider a family of stationary random probability fields[1] for the compliance $\mathcal{C}$. They are completely characterized by the marginal distribution and two parameters L_c and α which fully characterize the covariance structure. To describe this family we use an auxiliary mean zero and unit stationary Gaussian field $G(x,\omega)$ that has covariance function

$$Cov[G](x-y) = \mathbb{E}[G(x)G(y)] = \rho_G(x-y) = \rho\left(\frac{x-y}{L_c}\right) = e^{-\left(\frac{|x-y|}{L_c}\right)^{\alpha}}. \quad (1)$$

The compliance model is then a transformation of the auxiliary field to match the desired marginal distribution, i.e.

$$\mathcal{C} = \mathcal{C}(x,\omega) = \mathbb{E}[\mathcal{C}] + std[\mathcal{C}]F^{-1} \circ \Phi\left(G(x,\omega)\right). \quad (2)$$

Here Φ is the cumulative distribution of standard normal random variable, while the function F is related to the marginal distribution of the compliance $\mathcal{C}$. In fact F is the marginal cumulative distribution of the normalized field

$$\mathcal{Z} = \frac{\mathcal{C} - \mathbb{E}[\mathcal{C}]}{std[\mathcal{C}]}.$$

13 Calibration problem

The calibration step uses the data from the calibration experiments, which are pointwise compliance measurements and the elongation of the calibration bar.

13.1 Fitting the compliance with parametric and non-parametric marginal distributions

We consider four parametric distributions to fit the compliance experimental data, namely: uniform, normal, lognormal and inverse lognormal, as well as two non-parametric distributions.

a) *Parametric distributions.* In all cases the parameters of the distributions are chosen to match the first two sample moments of the compliance data. We denote by

$$\mathcal{A}[\mathcal{C};N] = \frac{1}{N}\sum_{j=1}^{N}\mathcal{C}(\omega_j) \quad (3)$$

the sample mean and by

$$\mathcal{S}[\mathcal{C};N] = \sqrt{\frac{1}{N-1}\sum_{j=1}^{N}(\mathcal{C}(\omega_j) - \mathcal{A}[\mathcal{C};N])^2} \quad (4)$$

[1]For the reader unfamiliar with these concepts, classical references on probability and Bayesian statistics, elementary and more advances, are for instance [McD04, Jay03, Lee04, BT73]

Case	$\mathcal{A}[\mathcal{C}; N_c] \times 10^{11}$	$\mathcal{S}[\mathcal{C}; N_c] \times 10^{11}$
1, $N_c = 5$	8.34	1.12
2, $N_c = 20$	8.68	0.64
3, $N_c = 30$	8.65	0.61

Table 6: Mean and standard deviation of the pointwise compliance for the three groups of samples

the sample standard deviation.

From the experimental data we get the mean and the standard deviation for the three groups of measurements, shown in Table 6. Using these data we get the parameters of the distributions:

1. Uniform distribution. Here we fit $\mathcal{C}(0, \omega) \sim U(A, B)$ and A, B are given in the Table 7

Case	$A \times 10^{11}$	$B \times 10^{11}$
1, $N_c = 5$	6.41	10.28
2, $N_c = 20$	7.58	9.78
3, $N_c = 30$	7.59	9.70

Table 7: The parameters A and B for the uniform distribution.

2. Normal distribution. Here we fit $\mathcal{C}(0, \omega) \sim N(\mu_G, \sigma_G^2)$. The parameters and given in Table 8.

Case	$\mu_G \times 10^{11}$	$\sigma_G \times 10^{11}$
1, $N_c = 5$	8.34	1.12
2, $N_c = 20$	8.68	0.64
3, $N_c = 30$	8.65	0.61

Table 8: The parameters μ_G and σ_G for the normal distribution.

3. Inverse log normal distribution. Here we fit $\mathcal{C}(0, \omega) = \frac{1}{E_{ILG} + \exp(N(\mu_{ILG}, \sigma_{ILG}^2))}$. The offset constant E_{ILG} is chosen to get positive minimal value of the modulus of elasticity E. The parameters μ_{ILG}, σ_{ILG} and E_{ILG} are given in Table 9.
4. Log-normal distribution. Here we fit $\mathcal{C}(0, \omega) \sim \exp(N(\mu_{LG}, \sigma_{LG}^2))$. The parameters μ_{LG} and $\sigma_{LG}$are given in Table 10.

Case	μ_{ILG}	σ_{ILG}	$E_{ILG} \times 10^{-10}$
1, $N_c = 5$	20.77	1.11	1.05
2, $N_c = 20$	20.77	0.617	1.03
3, $N_c = 30$	20.82	0.612	1.03

Table 9: The parameters of the inverse log normal distribution.

Case	μ_{LG}	σ_{LG}
1, $N_c = 5$	-23.21	0.14
2, $N_c = 20$	-23.17	0.078
3, $N_c = 30$	-23.17	0.074

Table 10: The parameters of the log normal distribution.

Fig. 8 shows the fitted cumulative distributions, together with the empirical one, for the three groups of measurements.

b) *Nonparametric distributions.* Here we are using the kernel density function based on two kernels, the Gaussian and the Epanetchnikov (for details see [BNT06]). The cumulative distribution based on the kernel density function, together with the empirical one, for the three groups of measurements are shown in Fig. 9.

Since the Gaussian and Epanetchnikov density functions are leading practically to the same results, only the former will be considered below.

13.2 Ranking the models for the marginal distribution

In Section 13.1 we constructed various models of the marginal distribution. We have to rank them with respect to the goal of the analysis. We use the discrepancy theory and the Kullback-Leibler measure (for details see [BNT06, KL51]) and use bootstrapping (see next section) to estimate the distribution of such discrepancies. We rank then all considered parametric and nonparametric marginal distributions according to the mean or median of the bootstrapped discrepancy distribution. Lower values characterize smaller discrepancy and hence higher quality of the model. The values are given in Table 11.

The uniform distribution leads to infinite discrepancy (for details see [BNT06]). We see that the inverse log-normal is the best, but the kernel density distribution is close. Note that the kernel density distribution is not assuming any particular form of probability density function (pdf) and is robust.

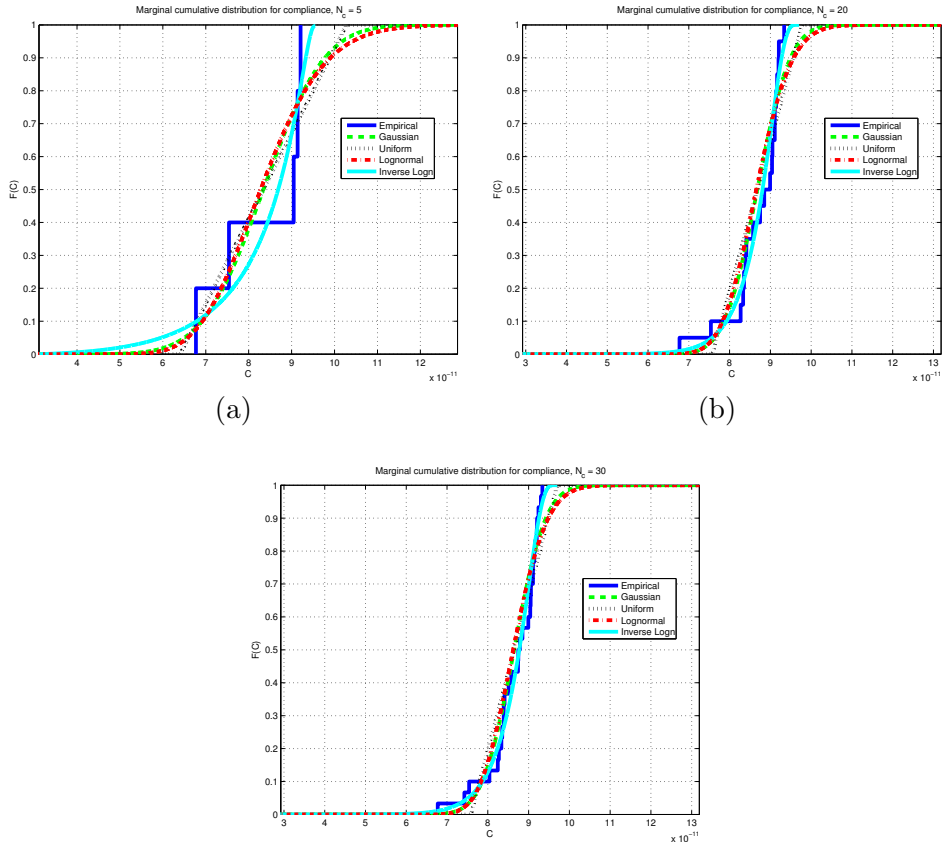


Figure 8: The fitted and the empirical cumulative distributions for the three groups of experiments: (a) $N_c = 5$, (b) $N_c = 20$, (c) $N_c = 30$.

13.3 Fitting the correlation length L_c

The covariance structure in (1) depends on two parameters, namely, α and the correlation length L_c. We will consider three cases: a) fully correlated random field ($\alpha = 0$, $L_c > 0$), b) partially correlated random field ($\alpha = 2$, $L_c > 0$) and c) perfectly uncorrelated field ($\alpha > 0$, $L_c = 0$). These choices were selected as the extreme cases although other intermediate cases could be considered as well. To get feeling on the two extreme choices we show in Fig. 10 the raw data for the compliance and elongation together with the relation between these data based on the models from the two extreme cases.

For $\alpha = 0$ the compliance is constant over the entire length of the sample and the relation between the compliance in the middle if the sample $C(L/2)$ and the elongation δL is linear.

In the uncorrelated case the elongation is independent of the compliance value $C(L/2)$. It is necessary to determine carefully this value from the raw

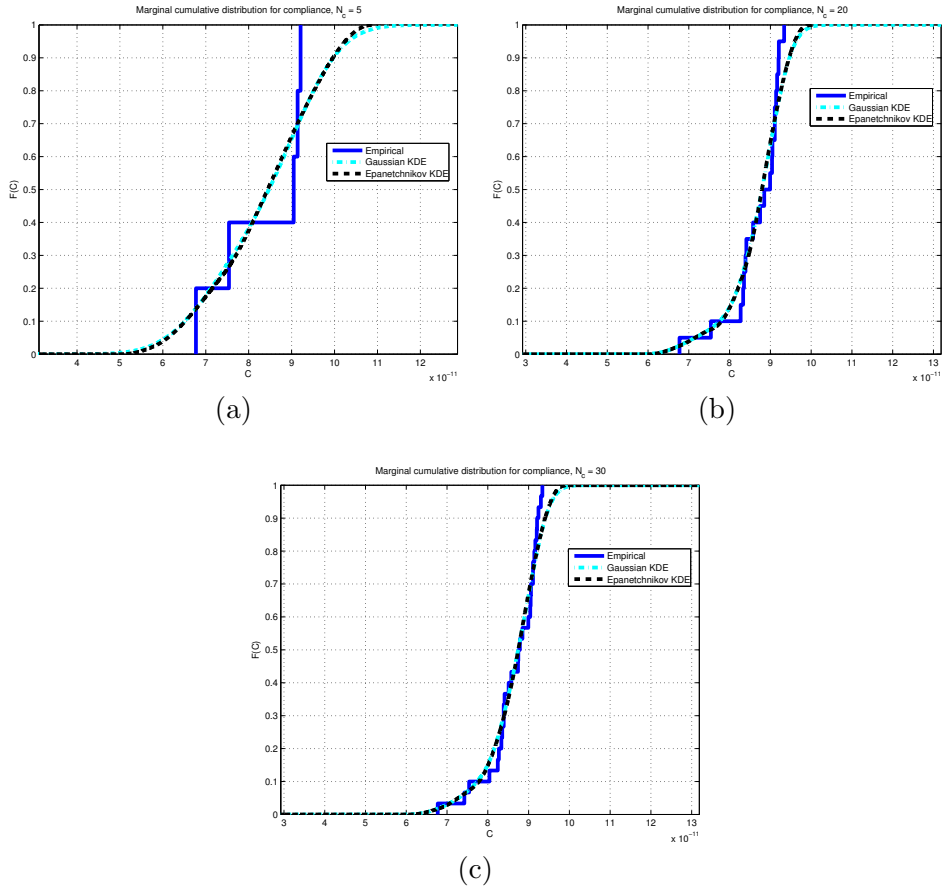


Figure 9: The cumulative distribution based on the kernel approach and the empirical distribution, for the three groups of the measurements: (a) $N_c = 5$, (b) $N_c = 20$, (c) $N_c = 30$.

data for the three groups of experiments. (For detail see [BNT06]). We see that the fully correlated case gives very bad results, while the uncorrelated case is reasonable.

In the case $\alpha = 2$ the correlation length L_c has to be determined. We observe that formulas (1) and (2), allow to determine the cross-covariance between $C(L/2)$ and the elongation δL, as well as the variance of δL, as a function of L_c. The same quantities (cross-covariance and variance) can also be determined from the experimental data. The correlation length L_c is then determined by least square so that the sample variance and cross-covariance are best fitted. The computed correlation length is given, for the three cases, in Table 12. The computed correlation length L_c was practically independent of the marginal distribution model.

	$N_c = 5$		$N_c = 20$		$N_c = 30$		Rank
Marginal	Median	Mean	Median	Mean	Median	Mean	(Median)
Uniform	+Inf	+Inf	+Inf	+Inf	+Inf	+Inf	(5)
Normal	1.38	+Inf	1.43	1.48	1.43	1.45	(3)
LogNormal	1.97	1.96	2.31	2.29	2.32	2.31	(4)
Inverse LN	**1.04**	**1.26**	**1.19**	**1.20**	**1.26**	**1.27**	(1)
KDE	1.35	+Inf	1.28	1.36	1.32	1.35	(2)

Table 11: Ranking of various models by the mean and the median of the bootstrapped discrepancy.

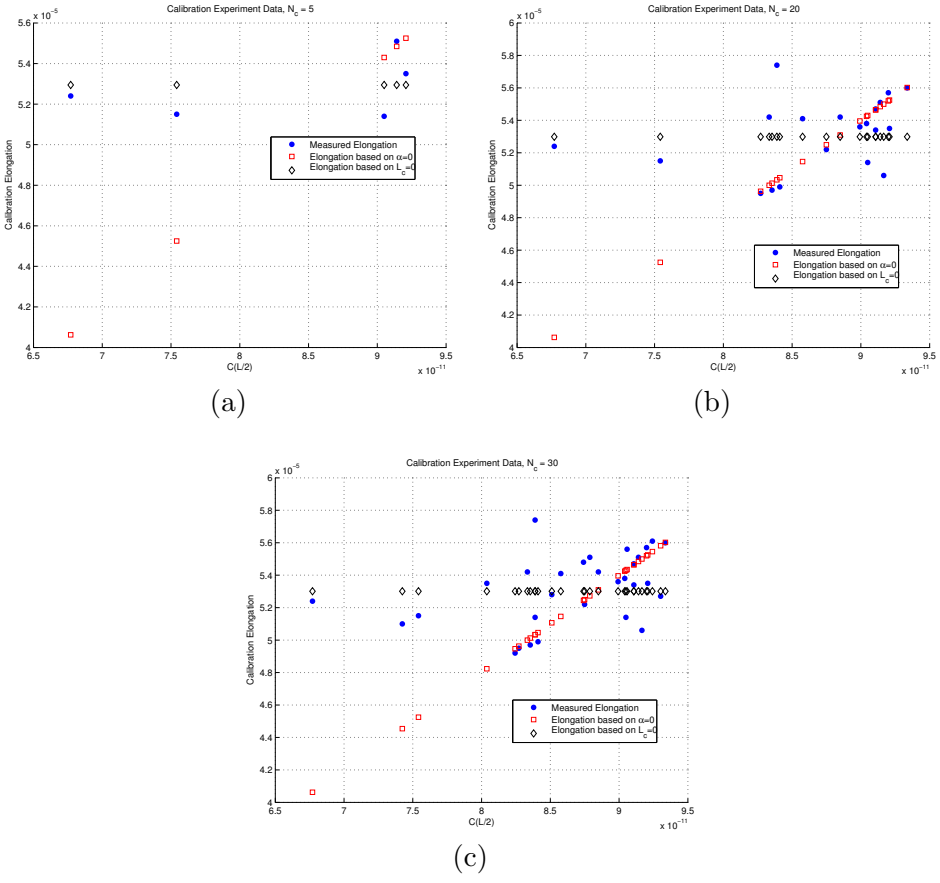


Figure 10: Elongation and point compliance data from the calibration experiment as well as the two extreme models, for the three groups of calibration experiments: (a) $N_c = 5$, (b) $N_c = 20$, (c) $N_c = 30$

Case	$L_c\,(m)$
1, $N_c = 5$	0.009
2, $N_c = 20$	0.030
3, $N_c = 30$	0.035

Table 12: The correlation length L_c determined from three sets of experimental data.

13.4 Variability in the fitted data

The variability of the parameters of the compliance models is very essential to establish the confidence in the model, especially in presence of a small set of experimental data. It is done by the boot-strapping approach [Efr82, ET93] as follows. The joint probability density function of $C(L/2)$ and δL is reconstructed by the Gaussian kernel density approach. The reconstruction is shown in Fig.11.

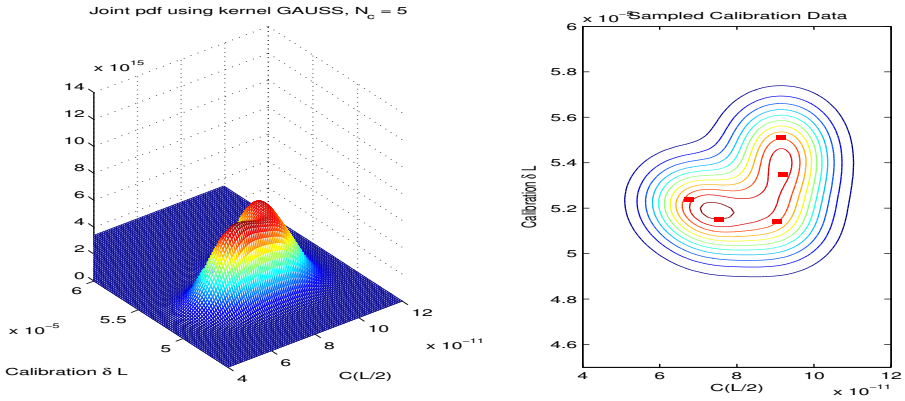
We draw samples of size N_c from the reconstructed pdf, and fit the parameters of the compliance model, as described in the previous sections, for each bootstrapped sample. We repeat the procedure $B = 1000$ times and get a distribution for the fitted parameters. In Fig. 12 we show the bootstrapped samples and the original experimental data.

Similarly it is possible to reconstruct the marginal distribution for $C(L/2)$ separately (or use the data from the joint probability). Then the bootstrapped cumulative distribution of the pointwise compliance can be computed (see Fig. 12 right). We see how the width of the distribution function decreases with increasing number of experiments. This width is related to the confidence in the cumulative distribution function.

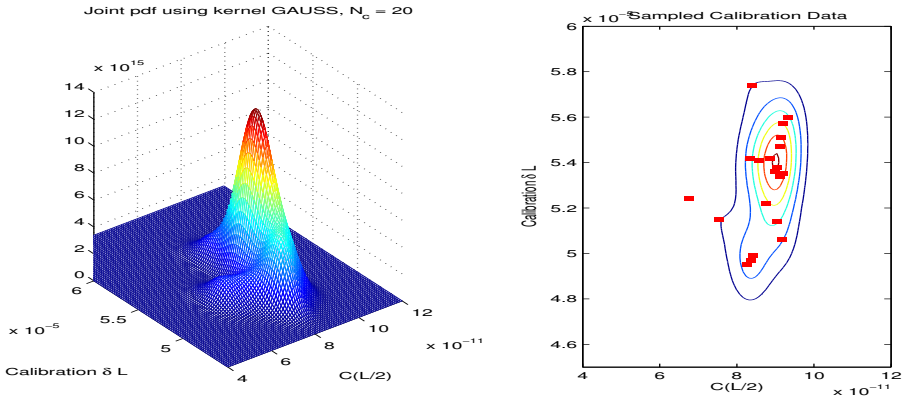
In Fig. 13 we show the bootstrapped sample compliance variance versus the compliance mean (left) and the sample compliance correlation length L_c versus the compliance mean. These figures show the confidence of the mean, variance and correlation length.

13.5 On the model and the calibration data

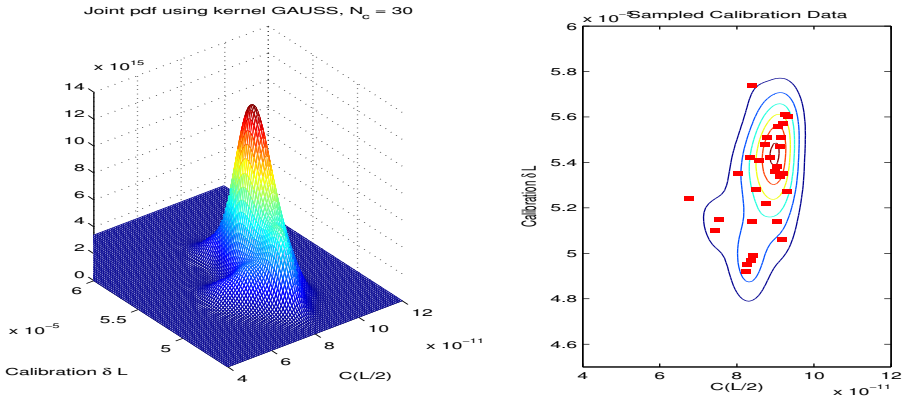
We considered a few models based on different marginal probabilities and covariance form (1). We concluded that two marginal distributions, the inverse log normal and the kernel density function lead to the best results. It was seen that the partially correlation model gives very reasonable results while the fully correlated model is inappropriate. The perfectly uncorrelated model gives practically acceptable results. It was clear that various models should be considered and ranked. The small number of experimental data significantly influences the approach. The model we consider is a stochastic model, which already characterizes the aleatory uncertainty. Nevertheless, we still have uncertainties of



(a)

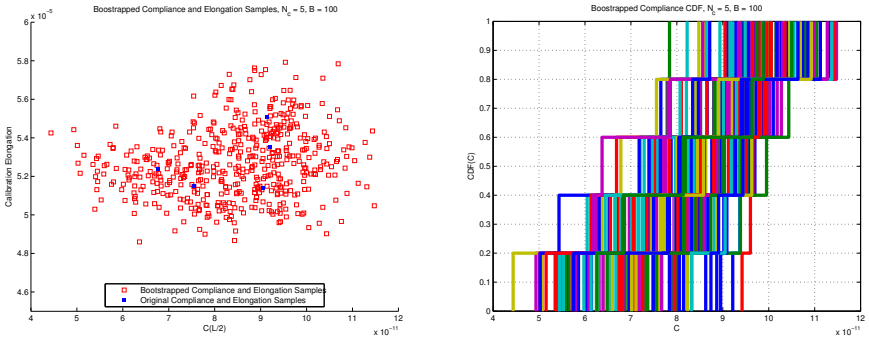


(b)

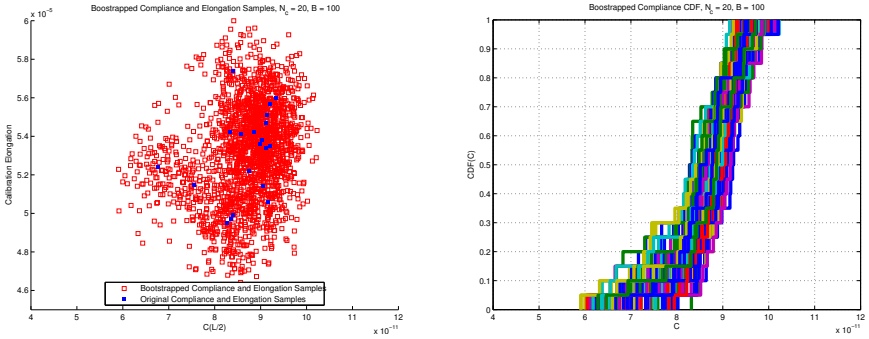


(c)

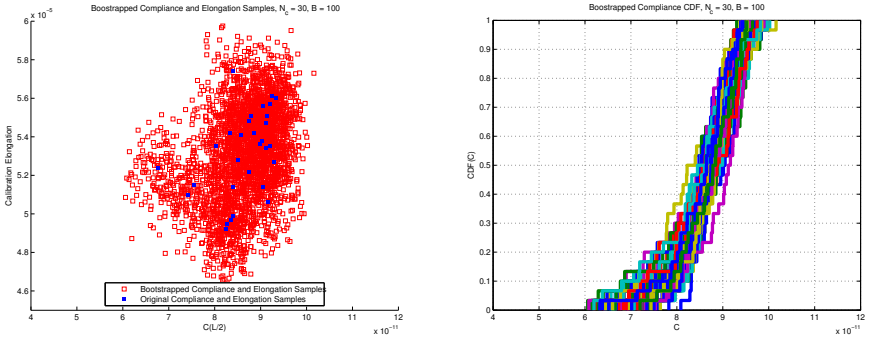
Figure 11: The reconstructed joint probability density from the calibration measurements, for the three groups: (a) $N_c = 5$, (b) $N_c = 20$, (c) $N_c = 30$. Right side shows the level lines. The squares are the original experimental data.



(a)

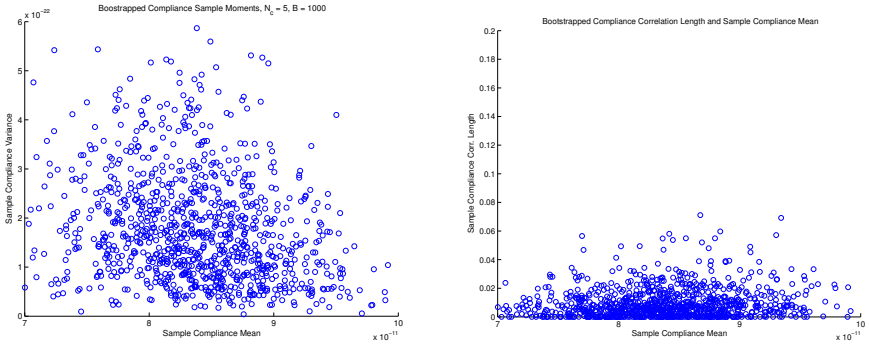


(b)

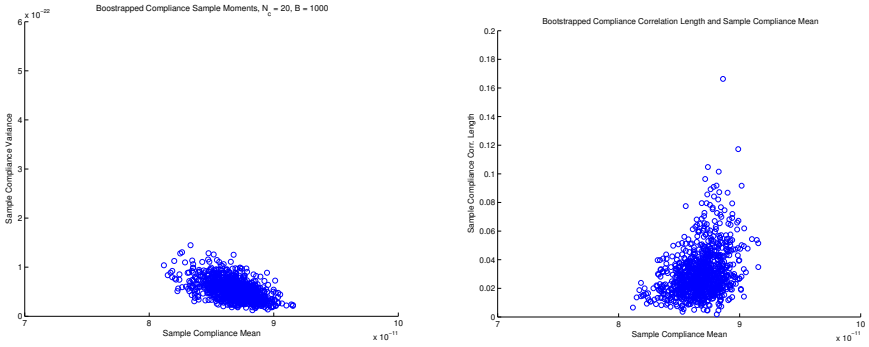


(c)

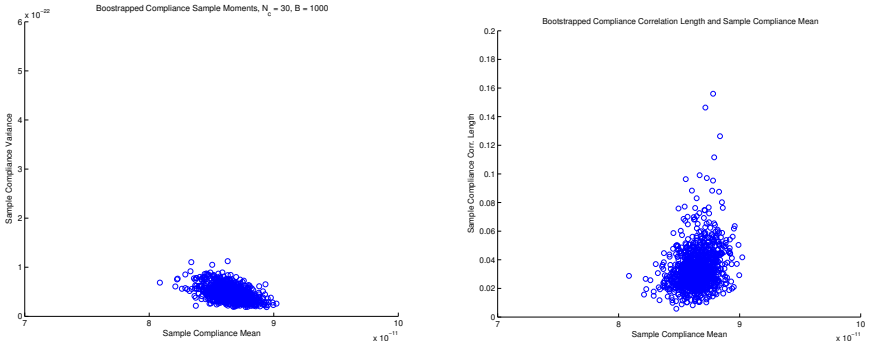
Figure 12: On the left: 100 bootstrapped samples from the reconstructed joint pdf of $C(L/2)$ and δL, for the three groups of experiments: (a) $N_c = 5$, (b) $N_c = 10$, (c) $N_c = 30$. On the right: bootstrapped empirical cumulative distribution of the compliance.



(a)



(b)



(c)

Figure 13: The bootstrapped compliance variance versus compliance mean (left) and correlation length versus compliance means (right), for the three groups of experiments: (a) $N_c = 5$, (b) $N_c = 20$, (c) $N_c = 30$.

epistemic type stemming from the small number of experimental data. The statistical analysis addressed in this section relates to the quantification of the uncertainties as discussed in general in section 5 and particularly in Section 10. After finishing the calibration, we proceed to the validation phase.

14 The Validation process

The goal of the validation is to assess reliability of the calibrated model. There is a much smaller number of validation experiments than for calibration. The process which is used in this paper has three parts:

a) Using the calibrated model, the probability field of the validation elongation is constructed, typically by the Monte Carlo method;

b) Using the available validation measurements a Bayesian update of the compliance model parameters is computed. A noninformative prior (see [Tan96, Jef98]) is used. The updating is in some cases computationally intensive. The updated model is assumed to be much more accurate then the calibrated one.

c) Using both the calibrated and updated model, the probability distribution of the quantity of interest (in the prediction problem) is constructed. A notion of the distance between these two probability distributions is defined; this distance is used as the metric. The reason for it is that the metric has to be closely related to the quantity of interest, i.e. the goal of the analysis. The distance is also used for the definition of the rejection criterion. Based on it and the given tolerance the calibrated model could be rejected

14.1 The distance between two cumulative distributions and the rejection criterion

Definition 1 (Horizontal distance between CDFs) *Let $F, G : \mathbb{R} \to [0,1]$ be two cumulative distributions and $\mathcal{I}_\varepsilon(G)$ the $\varepsilon-$level*

$$\mathcal{I}_\varepsilon(G) \equiv \{x \in \mathbb{R} \mid \frac{\varepsilon}{2} \leq G(x) \leq 1 - \frac{\varepsilon}{2}\}.$$

We define the distance between F and G as

$$d_\varepsilon(F,G) := \max_{x \in \mathcal{I}_\varepsilon(G)} |F^{-1} \circ G(x) - x| \tag{5}$$

The distance is well suited for the quantity of interest which has to be predicted, i.e. the probability that the displacement in the point P_m is not exceeding $3mm$.

Fig. 14 gives graphical interpretation of the distance.

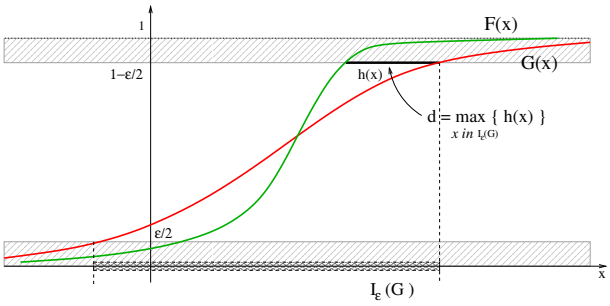


Figure 14: The graphical interpretation of the distance, which is related to the metric used in the validation.

The idea of the distance is the following. Assume that Y is the true value of the quantity of interest and Y^{cal} resp Y^{up} are the quantities predicted by the calibrated, respectively updated, model. Then we assume that

$$d_\varepsilon(F_Y, F_{Y^{cal}}) \approx d_\varepsilon(F_{Y^{up}}, F_{Y^{cal}}). \tag{6}$$

The value of the parameter ε is more or less arbitrary. We selected $\varepsilon = 0.005$. The reason is to avoid the reliance on the tail which is unstable and practically not significant if it is sufficiently small as in our case. Then the neighborhood of size +/- ε is used as the probability box which characterizes the reliability of the model. This leads directly to the definition of the metric and the criterion based on a given tolerance *tol*:

Definition 2 (Rejection criterion) *Given a tolerance level* tol *we reject the model* Y^{cal} *if*

$$d_\varepsilon(F_{Y^{up}}, F_{Y^{cal}}) \geq tol \times 3mm \tag{7}$$

The value of the tolerance could be different and has to be related to the decision based on the computed data. Selecting a larger tolerance, the cruder and simpler model will not be rejected but the reliability of the prediction will be smaller. This has to be taken into consideration when the decision (for which the computation is the basis) is made. If the tolerance is small, then possibly all models will be rejected when only small number of data is available. The models not rejected could be much more complex and computationally more expensive. In what follows we use $tol = 0.10$. Let us mention that rejection depends not only on the tolerance but also on the value of the parameter ε and the selection of both depends on the purpose of the computation.

14.2 The Bayesian updating of the parameters

For every model we compute the updated parameters using the noninformative prior. Here we will report only the mean and standard deviation of updated marginal distribution, for the three models:

(a) Full correlation, $\alpha = 0$, $L_c > 0$. In Table 13 we give the ratio of the means and standard deviations (std) between the updated and the calibrated models.

	Uniform		Normal		Inverse LN	
Case	$\frac{\mu_U^{up}}{\mu_U^{cal}}$	$\frac{\sigma_U^{up}}{\sigma_U^{cal}}$	$\frac{\mu_G^{up}}{\mu_G^{cal}}$	$\frac{\sigma_G^{up}}{\sigma_G^{cal}}$	$\frac{\mathbb{E}[\mathcal{C}(0)]^{up}}{\mathbb{E}[\mathcal{C}(0)]^{cal}}$	$\frac{std[\mathcal{C}(0)]^{up}}{std[\mathcal{C}(0)]^{cal}}$
1, $N_v = 2$	1.02	0.055	1.02	0.1	1.01	0.1
2, $N_v = 4$	0.98	0.14	0.98	0.2	0.97	0.2
3, $N_v = 10$	0.99	0.26	0.99	0.3	0.98	0.3

Table 13: The ratio of the calibrated and updated means and std for uniform, normal and inverse log normal marginal distribution and full correlation.

(b) Partial correlation, $\alpha = 2$, and calibrated value of L_c. In this case, we do not update the covariance length L_c because $L_v \gg L_c$ and the influence of specific value of L_c is very small. In Table 14 we show the ratios for the partial correlation model, for uniform, normal and inverse log normal marginal distributions

	Uniform		Normal		Inverse LN	
Case	$\frac{\mu_U^{up}}{\mu_U^{cal}}$	$\frac{\sigma_U^{up}}{\sigma_U^{cal}}$	$\frac{\mu_G^{up}}{\mu_G^{cal}}$	$\frac{\sigma_G^{up}}{\sigma_G^{cal}}$	$\frac{\mathbb{E}[\mathcal{C}(0)]^{up}}{\mathbb{E}[\mathcal{C}(0)]^{cal}}$	$\frac{std[\mathcal{C}(0)]^{up}}{std[\mathcal{C}(0)]^{cal}}$
1, $N_v = 2$	1.01	0.57	1.01	0.56	1.02	0.57
2, $N_v = 4$	0.98	0.81	0.98	0.80	0.98	0.80
3, $N_v = 10$	0.99	0.98	0.99	1.00	0.98	1.00

Table 14: The ratio of the calibrated and updated means and std for the uniform, normal and inverse log normal marginal distributions for the partial correlation model.

(c) Perfect non-correlation, $L_c = 0$. This case has to be handled differently because the problem become deterministic with the material described by the effective modulus E_{eff} only. The predicted elongation has $std = 0$ and depends only on the mean value μ independently of the marginal distribution. Hence only the mean can be updated. Yet this updating has a drawback because its variation cannot be updated and hence the assumption that the updated model is accurate does not hold. Therefore as a more accurate model we use the normal distribution with the updated mean and the standard deviation computed from the measured data.

14.3 Distance computation and model rejection

Once we have updated the models we can compute the cumulative distribution of the predicted quantity of interest by the calibrated and the updated models. Then we compute the distance between both distributions. Table 15 shows the ratio of the distance and the critical displacement $3mm$ for the three groups of the validation experiments. In boldface are the rejected models for a tolerance $tol = 0.10$. We see that for the tolerance $tol = 0.10$ the fully correlated model

	Case 1	Case 2	Case 3
$(N_c, N_v) =$	$(5, 2)$	$(20, 4)$	$(30, 10)$
	(a) *Fully correlated model* $\alpha = 0$		
Uniform	**0.25**	**0.14**	**0.12**
Normal	**0.25**	**0.13**	**0.11**
Inverse LN	**0.25**	**0.17**	**0.12**
	(b) *Partially correlated model* $\alpha = 2$		
Uniform	0.02	0.04	0.025
Normal	0.03	0.035	0.025
Inverse LN	0.025	0.035	0.03
	(c) *Perfectly uncorrelated model* $L_c = 0$		
Effective model	0.07	0.08	0.09

Table 15: The ratio of the horizontal distances and the critical prediction displacement.

has to be rejected for all marginal distributions although its reliability slightly grows with the increase of the number of experiments. Of course this model is the simplest one. The partial correlation model leads to the best results for all marginal distributions. The completely uncorrelated one still leads to a reasonable accuracy and it is computationally simpler than the partially correlated model. Hence we do reject the completely correlated model and keep the partially correlated and uncorrelated ones. Note that the ratios in Table 15 are not in all cases decreasing when increasing the number of experiments. This is caused by the influence of the small number of experiments.

To further illustrate the results we show, on the left side of Fig. 15 the cumulative distribution of the elongation L_v, empirical, predicted by the calibrated

model and by the updated one, for the uniform marginal and fully correlated model. On the right side we show the cumulative distribution of the quantity of interest (prediction) and the bounds based on the computed distance. Note that on the left the cumulative distribution of the elongation L_v has been computed analytically, while on the right the cumulative distribution has been computed by the Monte Carlo Method (although it could also have been computed analytically as well). In Fig. 16 we show analogous results for the partially correlated model with normal marginal distribution.

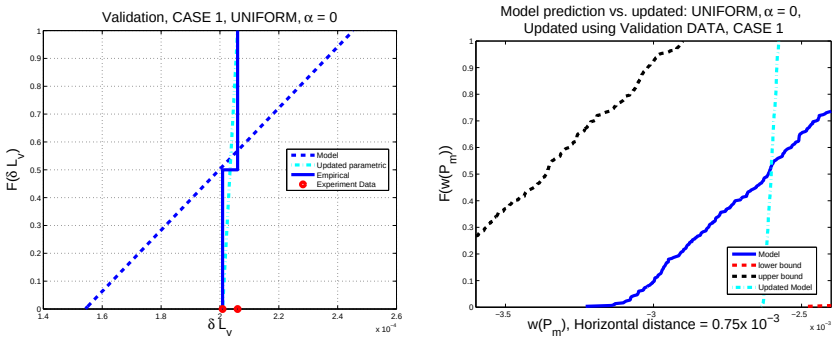
15 The accreditation

The accreditation process is analogous to the validation one. We have to take here into account that only a very small number of experimental data is available. Using the calibrated model we compute the displacement of the midpoint Q of bar 1 in the accreditation frame, and by the Bayesian approach we update the mean and the std of the compliance model using the available accreditation measured data. In the case when only one experiment is available, only the mean is updated and. Then we compute the cumulative distribution of the quantity of interest in the prediction problem using the calibrated and updated model and compute the distance of these two distributions. In Table 16 we give the ratio of the distance and the critical value. Because the fully correlated model was rejected on the validation level we do not address it in the accreditation phase. From the table we see that with the tolerance $tol = 0.1$ we do not reject the

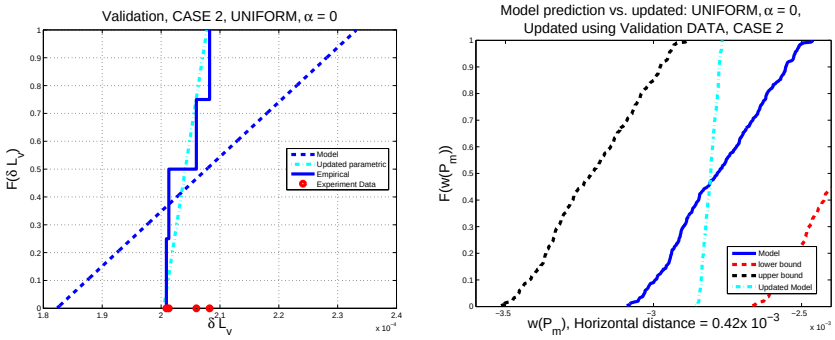
$(N_c, N_a) =$	Case 1 $(5,1)$	Case 2 $(20,1)$	Case 3 $(30,2)$
	(a) *Partially correlated model* $\alpha = 2$		
Uniform	0.015	0.028	0.037
Normal	0.015	0.028	0.037
Inverse LN	0.015	0.026	0.033
	(b) *Perfectly uncorrelated model* $L_c = 0$		
Effective model	0.072	0.072	0.1

Table 16: The ratio of the distance and the critical value, in the accreditation procedure, for the partially correlated models and the perfectly uncorrelated model.

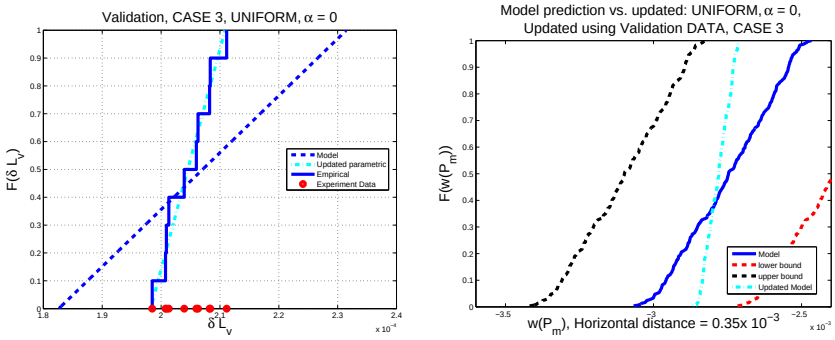
models, with the exception of the perfectly uncorrelated model in the case 3.



Case 1: $N_c = 5,\ N_v = 2$

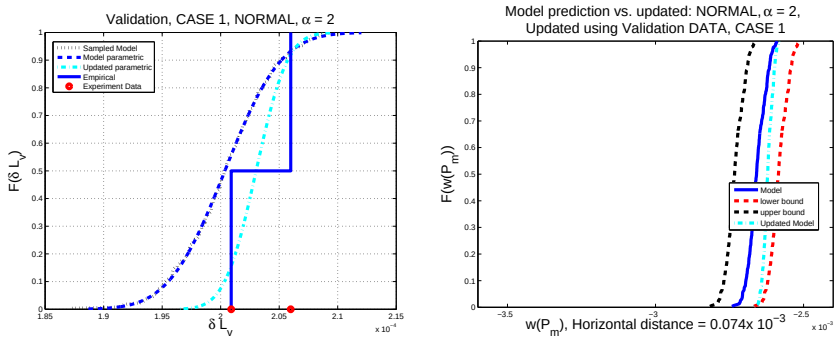


Case 2: $N_c = 20,\ N_v = 4$

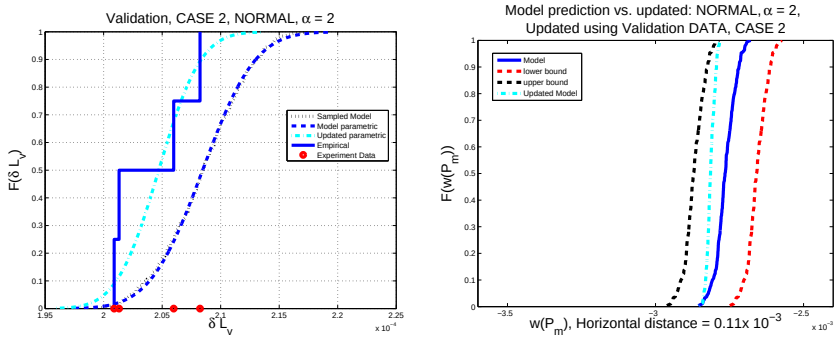


Case 3: $N_c = 30,\ N_v = 10$

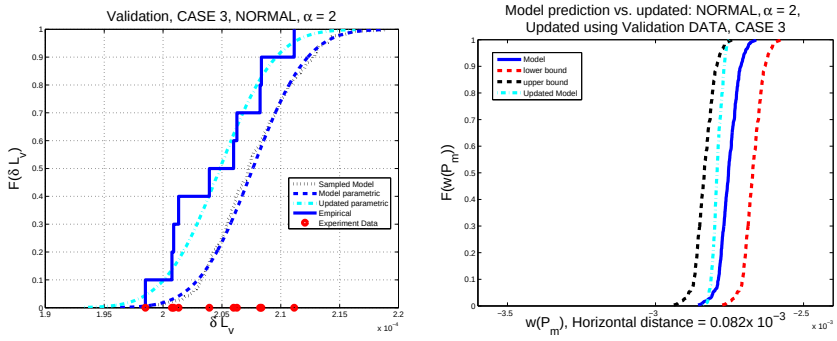
Figure 15: The uniform marginal distribution and full correlation. On the left is the cumulative distribution of the elongation L_v, empirical, predicted by the calibrated model and by the updated one. On the right is the cumulative distribution of the quantity of interest with the bounds based on the distance. The bounds are creating the probability box.



Case 1: $N_c = 5,\ N_v = 2$



Case 2: $N_c = 20,\ N_v = 4$



Case 3: $N_c = 30,\ N_v = 10$

Figure 16: The partially correlated model with normal marginal distribution. On the left is the cumulative distribution of the elongation L_v, empirical, predicted by the calibrated model and by the updated one. On the right is the probability box for the quantity of interest, based on the computed distance.

16 The prediction

As we have said above the prediction is based on the calibrated model taking into account also the variability in the fitted data, estimated by bootstrapping, especially in the case of a small number of calibration experiments. The distances computed in the validation and accreditation processes are used for determining bounds in the predicted failure. The goal is to approximate $P(w(P_m) > 3mm)$ where $w(P_m)$ is the displacement in the midpoint P_m of the bar 4 in the frame. To properly account for the variability in the prediction due to the small amount of calibration information, we consider the parameters defining the compliance model (2) (hereafter denoted by Θ) as random variables and sample them by bootstrapping, using the procedure described in Section 13.4. If $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{P})$ is the probability space for Θ and $(\Omega, \mathcal{G}, Q)$ is the probability space for the model prediction displacement for a given Θ, then

$$P(|w(P_m, \Theta)| \geq 3(mm)) = \int_{\tilde{\Omega}} \int_{\Omega} \mathbf{1}_{\{|w(P_m, \Theta(\tilde{\omega}), \omega)| \geq 3(mm)\}} dQ(\omega) d\tilde{P}(\tilde{\omega}). \qquad (8)$$

Given two natural numbers B and M we hierarchically sample first B times the model parameters Θ by bootstrapping, then, for each sampled Θ, we sample $w(P_m, \Theta)$ M times to generate $B \times M$ corresponding bootstrap samples. This is computationally intensive and various simplification described in [BNT06] have been made. This procedure is important when only a small number of calibration experiments is available. We note that also for a large number of experiments the bootstrapping results will not coincide with the prediction based on the calibrated model. This is because the calibrated model is only approximate. In Table 17 we are giving the failure probability evaluated by the bootstrapping for various marginal probability distributions.

$(N_c, N_v, N_a) =$	Case 1 $(5, 2, 1)$	Case 2 $(20, 4, 1)$	Case 3 $(30, 10, 2)$
Uniform	2.0×10^{-2}	4.7×10^{-5}	5.6×10^{-8}
Normal	2.1×10^{-2}	5.8×10^{-5}	8.6×10^{-8}
Inverse LN	1.4×10^{-2}	4.6×10^{-5}	5.6×10^{-8}

Table 17: Failure probabilities for the prediction displacement $P(|w(P_m)| \geq 3mm)$, evaluated with the bootstrapping procedure for the partially correlated models.

We augment the results by the information we have obtained during the validation and accreditation phases to get bounds. For this, instead of providing as failure probability the quantity $P(|w(P_m)| > 3mm)$ we use $P(|w(P_m)| > 3(1 \pm d_\varepsilon)mm)$, where d_ε is the maximum of the distances measured in the validation

and the accreditation phases. In Table 18 we show upper bounds on the failure probability, evaluated by combining the bootstrapping and the distances in the validation and accreditation procedure.

$(N_c, N_v, N_a) =$	Case 1 $(5, 2, 1)$	Case 2 $(20, 4, 1)$	Case 3 $(30, 10, 2)$
Uniform	0.18	1.8×10^{-2}	3.0×10^{-3}
Normal	0.19	2.1×10^{-2}	2.1×10^{-3}
Inverse LN	0.19	1.7×10^{-2}	1.4×10^{-3}

Table 18: The probability of failure bounds obtained by combining the bootstrapping procedure with the distances measured in the validation and accreditation experiments, for various marginal probabilities and the partially correlated model.

In Fig. 17 we show the cumulative distribution of the prediction quantity computed by the calibrated model and by the bootstrapped one (which accounts also for the variability of the compliance model parameters) for the normal marginal distribution and partial correlation. In addition we show the bounds based on the distances in the validation and accreditation. In Fig. 18 we show analogous results for the perfectly uncorrelated model.

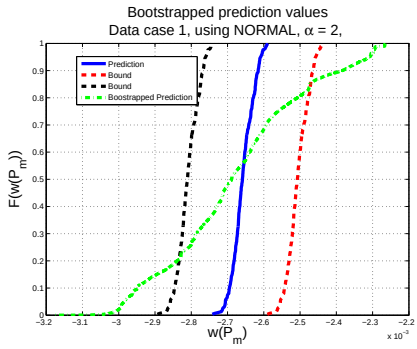
From Figures 17 and 18 we see that for the case of the smallest number of experimental data the bootstrapping is essential. For the higher number of experiments the influence of the "modeling error", quantified by distances measured in the validation and accreditation, is larger than the variability related to the number of experiments, estimated by bootstrapping.

Comparing Figs. 17 and 18 we see that the use of the bounds is essential for the confidence in the perfectly uncorrelated model which is less reliable than the partially correlated model.

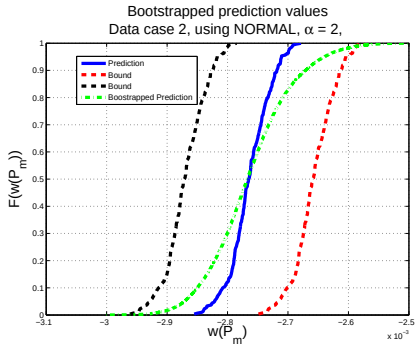
17 Conclusions

We have shown the general process of the prediction based on the calibration, validation and accreditation. We have seen the role of the number of experimental data. If the number is small then the accuracy of the prediction can be very unreliable and the decision has to take it properly into consideration.

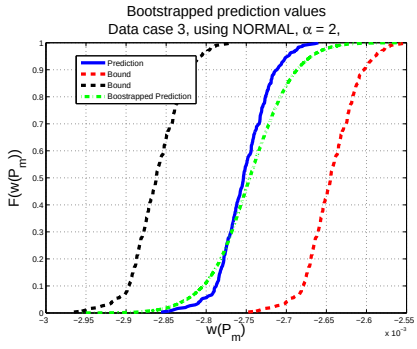
The validation analysis for the academic frame problem was based on the Bayesian probability. This approach may be costly in general, yet there are ways to perform the Bayesian update approximately, thus reducing the computational cost. Other techniques such as sensitivity analysis or worst-case scenario approach could be used as well.



Case 1: $N_c = 5,\ N_v = 2,\ N_a = 1$

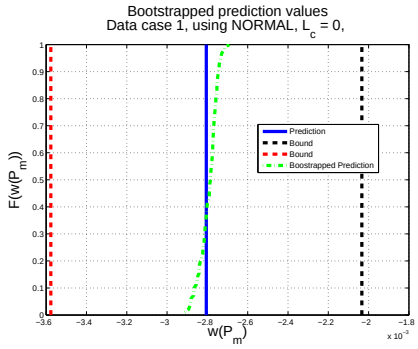


Case 2: $N_c = 20,\ N_v = 4,\ N_a = 1$

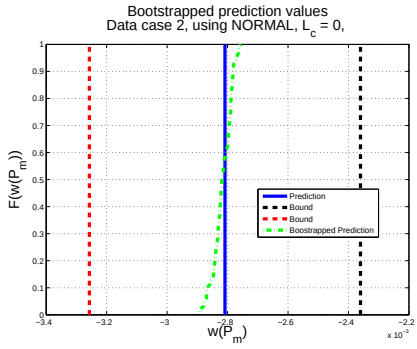


Case 3: $N_c = 30,\ N_v = 10,\ N_a = 2$

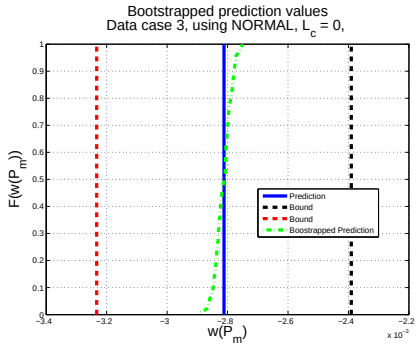
Figure 17: The cumulative distribution of the predicted quantity of interest obtained by the calibrated model (blue), the bootstrapped model (green) and the bounds. Partially correlated model with Normal marginal distribution.



Case 1: $N_c = 5,\ N_v = 2,\ N_a = 1$



Case 2: $N_c = 20,\ N_v = 4,\ N_a = 1$



Case 3: $N_c = 30,\ N_v = 10,\ N_a = 2$

Figure 18: The cumulative distribution of the predicted quantity of interest by the calibrated (blue) and the bootstrapped model (green) for the perfectly uncorrelated model.

In our problem we neglected the errors in the experiments. These of course have to be addressed in any analysis too. Many approaches to the problem of validation under uncertainties exist. In our academic problem the exact solution i.e. the probability of the failure is not available. It seems to be very important to develop various approaches and test them on problems where the exact solution is known (the manufactured problems). In the general case, as for instance the airplane design, the situation is much more complex and many theoretical and implementational issues are open. The problem of the adaptive modeling which leads to the best results with minimal cost is also essential.

Acknowledgments

We would like to thank Dr. Laurent Chambon (EADS-CCR) and Dr. Stéphane Guinard (EADS-CCR) for valuable comments.

References

[AIA98] Guide for verification and validation of computational fluid dynamics simulation. Technical Report AIAA G-077-1998, American Institute of Aeronautics and Astronautics, 1998.

[Bab61] I. Babuška. On randomized solutions of Laplace's equation. *Casopis Pest. Math.*, 86:269–276, 1961.

[BJLS93] I. Babuška, K. Jerina, Y. Li, and P. Smith. Quantitative assesment of the accuracy of constitutive laws for plasticity with an exposure on cyclic deformation. *Material Parameter Estimation for Modern Constitutive Equations, ASME*, MD-Vol 43/ AMD- Vol 168:116–169, 1993.

[BLT03] I. Babuška, K-M. Liu, and R. Tempone. Solving stochastic partial differential equations based on the experimental data. *Math. Models Methods Appl. Sci.*, 13(3):415–444, 2003. Dedicated to Jim Douglas, Jr. on the occasion of his 75th birthday.

[BNT05] I. Babuška, F. Nobile, and R. Tempone. Worst-case scenario analysis for elliptic problems with uncertainty. *Numer. Math.*, 101:185–219, 2005.

[BNT06] I. Babuška, F. Nobile, and R. Tempone. Model validation challange problem: Static frame prediction. sandia validation workshop. SANDIA Workshop, May 27-29,2006, 2006.

[BO04] I. Babuška and J-T. Oden. Verification and Validation in Comptuational Engineering and Science. Part I: Basic concepts. *Comp. Meth. Appl..Mech. Engrg.*, 193:4057–4061, 2004.

[BO05] I. Babuška and J-T. Oden. The reliability of computer predictions; can they be trusted. *Int. J. Num. Anal. Mode.*, 3, 2005.

[BS01] I. Babuška and T. Strouboulis. *The Finite Element Method and its Reliability*. Oxford University Press, 2001.

[BT73] G. E. P. Box and G. C. Tiao. *Bayesian inference in statistical analysis*. Addison-Wesley Publishing Co., Reading, Mass.-London-Don Mills, Ont., 1973. Addison-Wesley Series in Behavioral Science: Quantitative Methods.

[BTZ04] I. Babuška, R. Tempone, and G.E. Zouraris. Galerkin finite element approximations of stochastic elliptic partial differential equations. *SIAM J. Numer. Anal.*, 42(2):800–825 (electronic), 2004.

[BTZ05] I. Babuška, R. Tempone, and G.E. Zouraris. Solving elliptic boundary value problems with uncertain coefficients by the finite element method: the stochastic formulation. *Comput. Methods Appl. Mech. Engrg.*, 194(12-16):1251–1294, 2005.

[CF99] A.C. Cullen and H.C. Frey. *Probabilistic Techniques Exposure Assessment*. Plenum Press, 1999.

[Efr82] B. Efron. *The jackknife, the bootstrap and other resampling plans*, volume 38 of *CBMS-NSF Regional Conference Series in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, Pa., 1982.

[ET93] B. Efron and R. J. Tibshirani. *An introduction to the bootstrap*, volume 57 of *Monographs on Statistics and Applied Probability*. Chapman and Hall, New York, 1993.

[HCB04] J. Hlaváček, I. Chleboun, and I. Babuška. *Uncertain input data problems and the worst scenario method*. Elsevier, Amsterdam, 2004.

[Jay03] E. T. Jaynes. *Probability theory*. Cambridge University Press, Cambridge, 2003. The logic of science, Edited and with a foreword by G. Larry Bretthorst.

[Jef98] Harold Jeffreys. *Theory of probability*. Oxford Classic Texts in the Physical Sciences. The Clarendon Press Oxford University Press, New York, 1998. Reprint of the 1983 edition.

[KL51] S. Kullback and R. A. Leibler. On information and sufficiency. *Ann. Math. Statistics*, 22:79–86, 1951.

[KOG98] G.B. Kleindorfer, L. O'Neil, and R. Ganeshan. Validation in simulation; various positions in the philosophy of science. *Management Science*, 44:1087–1099, 1998.

[Lee04] P. M. Lee. *Bayesian statistics*. Arnold, London, third edition, 2004. An introduction.

[McD04] D. McDonald. *Elements of applied probability for engineering, mathematics and systems science*. World Scientific Publishing Co. Inc., River Edge, NJ, 2004.

[NF67] T.H. Naylor and J.M. Finger. Verification of computer simulation model. *Management Science*, 14:B92– B103, 1967.

[OT02] W.L. Oberkampf and T.G. Trucano. Verification and validation in computational fluid dynamics. *Progress in Aerospace Sciences*, 38:209–272, 2002.

[Pos04] D.E. Post. The coming crisis in computational sciences. Technical Report Rep.LA-UR-03-88-2004, Los Alamos National Laboratory, 2004.

[Roa98] P.J. Roache. *Verification and Validation in Computational Science and Engineering*. Hermosa, Albuquerque, New Mexico, 1998.

[Tan96] M. A. Tanner. *Tools for statistical inference*. Springer Series in Statistics. Springer-Verlag, New York, third edition, 1996. Methods for the exploration of posterior distributions and likelihood functions.