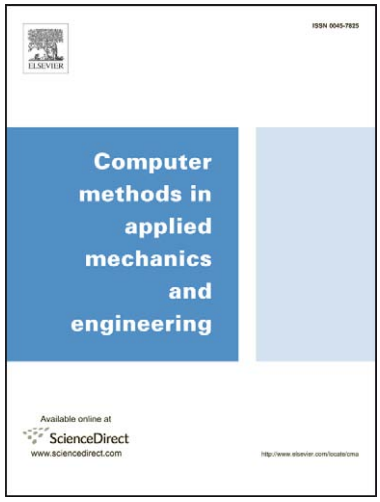


Rigorous Verification, Validation, Uncertainty Quantification and Certification through Concentration-of-Measure Inequalities

L.J. Lucas, H. Owhadi, M. Ortiz

PII: S0045-7825(08)00232-6
DOI: 10.1016/j.cma.2008.06.008
Reference: CMA 8700

To appear in: *Comput. Methods Appl. Mech. Engrg.*

Received Date: 11 August 2007
Revised Date: 27 May 2008
Accepted Date: 4 June 2008

Please cite this article as: L.J. Lucas, H. Owhadi, M. Ortiz, Rigorous Verification, Validation, Uncertainty Quantification and Certification through Concentration-of-Measure Inequalities, *Comput. Methods Appl. Mech. Engrg.* (2008), doi: 10.1016/j.cma.2008.06.008

Rigorous Verification, Validation, Uncertainty Quantification and Certification through Concentration-of-Measure Inequalities

L. J. Lucas, H. Owhadi and M. Ortiz
Graduate Aeronautical Laboratories
California Institute of Technology
Pasadena, CA 91125, USA

May 26, 2008

Abstract

We apply *concentration-of-measure* inequalities to the quantification of uncertainties in the performance of engineering systems. Specifically, we envision *uncertainty quantification* in the context of *certification*, i. e., as a tool for deciding whether a system is likely to perform safely and reliably within design specifications. We show that concentration-of-measure inequalities rigorously bound probabilities of failure and thus supply conservative certification criteria. In addition, they supply unambiguous quantitative definitions of terms such as margins, epistemic and aleatoric uncertainties, verification and validation measures, confidence factors, and others, as well as providing clear procedures for computing these quantities by means of concerted simulation and experimental campaigns. We also investigate numerically the tightness of concentration-of-measure inequalities with the aid of imploding and exploding ring examples. Our numerical tests establish the robustness and viability of concentration-of-measure inequalities as a basis for certification in that particular example of application.

1 Introduction

This paper is concerned with the application of *concentration-of-measure* inequalities to the quantification of uncertainties in the performance of engineering systems. Specifically, we envision *uncertainty quantification* (UQ) in the context of

certification, i. e., as a tool for deciding whether a system is likely to perform safely and reliably within design specifications.

The certification process is sometimes described in terms of quantification of margins and uncertainties (QMU) (cf, e. g., [1], [2], [3]). Thus, suppose that a system is required to perform at a certain level, or *threshold*, for its safe operation. The system is designed so that its performance under a worst-case scenario of potential operating conditions is somewhat higher than its threshold. A suitable measure M of the distance between these two levels constitutes the *performance margin*. However, because the systems of interest–and the conditions they operate in–are subject to statistical variation, performance measures are stochastic in nature. Therefore, the precise values of the expected performance level and its threshold are often uncertain. Sources of uncertainty include operation conditions such as loads and system characteristics such as geometry and material behavior. If U denotes a suitable measure of these uncertainties, the *confidence factor*

$$\mathrm{CF} = \frac{M}{U} \tag{1}$$

may be taken as a rational basis for certification. Thus, if CF is sufficiently larger than 1, the system may be regarded as safe and reliable and a candidate for certification.

A first obvious strategy is to attempt certification solely by means of experimental testing and statistical sampling. This approach is indeed the only approach possible when no *a priori* information is available about the behavior of the system. However, the number of tests required for purely empirical certification can be prohibitively high, especially in systems for which testing is expensive (cf Section 2.5), which renders the empirical approach unfeasible in practice. The problem is compounded when systems operate under extreme conditions outside the range of direct laboratory testing. Examples include large space structures and high-energy systems such as fusion reactors.

Under these conditions, *modeling* can provide a viable alternative to testing. Thus, complex systems can often be modeled through the application of sound physical and engineering principles and the resulting models can be used to accurately predict the performance of the system. In these cases, the central question concerns how the availability of a model can be exploited in order to achieve certification with the least number of tests. It is precisely in this potential reduction in the number of tests required for certification that the benefit of model-based certification lies. Evidently, the extent to which a model can reduce the amount of testing depends on the *predictiveness* of the model. The assessment of model predictiveness is accomplished through *verification and validation*, i. e., through a careful assessment of numerical and modeling errors.

Compelling as this picture is, the challenge of rendering it in *rigorous* and precise mathematical terms–and of developing a set of computational tools enabling its efficient implementation–is quite considerable. By rigorous UQ we specifically mean a set of mathematically provable inequalities that provide rigorous upper bounds for the probability of failure of the system. While verification and validation (V&V) and UQ have been the subject of extensive research in recent years (cf, e. g., [4], [1], [5], [2], [3]), a rigorous mathematical and computational framework of the type just described does not appear to had been devised to date. For instance, in the context of QMU rigorous (in the sense just described) definitions of M and U—and tractable means of computing them—are often left unspecified. The problem is compounded when system certification requires the quantification of uncertainties in a number of performance measure simultaneously. In such cases, *ad hoc* methods for aggregating uncertainties, e. g., by root-mean squares, are often used without justification. A clear connection between the confidence factor CF and the probability of failure of the system is also often lacking. In addition, precise quantitative measures of model predictiveness, the sequence of verification and validation tests required for computing such measures, and the precise manner in which models aid rigorous certification are also often left unspecified, which reduces the value of QMU and model-based certification to that of a compelling but imprecise and heuristic conceptual framework.

The purpose of this paper is to develop a rigorous theory of uncertainty quantification and certification in which the requisite probability-of-failure upper bounds are supplied by *concentration-of-measure* inequalities. Roughly speaking, the concentration-of-measure phenomenon is a consequence of the fact that functions of a large number of variables, i. e., functions in high-dimensional spaces, with small local oscillations with respect to each variable are nearly constant. Moreover, fluctuations can be controlled through elementary yet powerful and non-trivial quantitative inequalities called concentration-of-measure inequalities (cf [6] for a monograph, [7] for a survey). These tools have found broad application in functional analysis, complexity theory, probability and statistics. However, the application of concentration-of-measure inequalities to uncertainty quantification does not appear to have been attempted to date.

In this paper we show how concentration-of-measure inequalities supply: probability-of-failure upper bounds resulting in rigorous and conservative certification; precise definitions and quantitative measures of margins and uncertainties enabling such certification; a precise relation between the confidence factor and the probability of failure of the system; a rigorous framework for model-based certification, including precise quantitative measures of model predictiveness and the effect of *unknown unknowns*; and the sequence of verification and validation tests required for assessing model predictiveness. The manner in which concentration-of-measure

supplies these figures of merit is demonstrated in Section 2 by means of a sequence of representative scenarios of increasing complexity. Section 3 illustrates the present approach to QMU in a particular example of application, the implosion and explosion of a ring, using the simplest variety of concentration-of-measure inequalities. Finally, the outlook of the approach is discussed in Section 4. A brief account of the development of concentration-of-measure inequalities in the field of probability and a representative sampling that showcases the range of applicability of the inequalities is presented in Appendix A for the convenience of the reader.

2 Concentration-of-measure inequalities applied to uncertainty quantification and certification

The essential mathematical problem of QMU is as follows. Consider a system with output performance measures $Y : \Omega \to E_1 \times \cdots \times E_N$ on a given probability space $(\Omega, \mathcal{U}, \mathbb{P})$. Here and subsequently throughout this work we use standard notation of probability theory (cf, e. g., Chapter 2 of [8] for an introduction). In particular, $E_1, \ldots, E_N$ denote Euclidean spaces endowed with the standard metric and $\mathbb{P}$ is a probability measure. Suppose that the safe operation of the system requires that $Y \in A$ for some open *admissible set* $A \subset E_1 \times \cdots \times E_N$. Ideally, we would then like the support of the probability measure associated to Y to be contained within A, i. e.,

$$\mathbb{P}[Y \in A] = 1 \tag{2}$$

Systems satisfying this condition can be certified with complete certainty. However, this absolute guarantee of safe performance may be unattainable, e. g., if $\mathbb{P}$ lacks compact support, or prohibitively expensive. In these cases, we may relax condition (2) to

$$\mathbb{P}[Y \in A^c] \leq \epsilon \tag{3}$$

for some appropriate *certification tolerance* ϵ, where $A^c = E_1 \times \cdots \times E_N \backslash A$ is the *inadmissible set*. Inequality (3) expresses the requirement that the probability of system failure be acceptably small and gives rigorous mathematical expression to the QMU conceptual view of certification.

A *conservative* certification criterion is obtained when the probability of failure $\mathbb{P}[Y \in A^c]$ is bounded from above and the upper bound is verified to be below the certification tolerance ϵ. Evidently, for an upper bound to be useful it must be *tight*, i. e., it must be close to the actual probability of failure $\mathbb{P}[Y \in A^c]$. Therefore, the essential mathematical and computational challenge is to obtain tight upper bounds to the probability of failure of the system.

failure tolerance (ϵ)	number of tests (m)
1	0
10^{-1}	115
10^{-2}	23,025
10^{-3}	3,453,877
10^{-4}	460,517,018
10^{-5}	57,564,627,324
10^{-6}	6,907,755,278,982

Table I. Number of tests required for certification as a function of the probability-of-failure tolerance.

A first strategy that naturally suggests itself is to bound the probability of failure empirically, i. e., solely by means of experimental testing. Suppose that m tests are performed with outcomes $Y^1, \ldots, Y^m$. With these data we can associate the empirical probability measure

$$\mu_m = \frac{1}{m}\sum_{i=1}^{m} \delta_{Y^i} \tag{4}$$

Then, an application of Hoeffding's inequality [9] gives, with probability $1-\epsilon'$,

$$\mathbb{P}[Y \in A^c] \leq \mu_m[A^c] + \big(\frac{1}{2m}\log\frac{1}{\epsilon'}\big)^{\frac{1}{2}} \tag{5}$$

Hence, the inequality

$$\mu_m[A^c] + \big(\frac{1}{2m}\log\frac{1}{\epsilon'}\big)^{\frac{1}{2}} \leq \epsilon \tag{6}$$

supplies a conservative certification condition. We note that inequality (5) can be improved by using Chernoff's inequality instead of Hoeffding's inequality when the probability of failure $\mathbb{P}[Y \in A^c]$ is known. Since that probability of failure is unknown, it can't be used practice to give a bound on the necessary number of tests one has to use Hoeffding's inequality.

The certification criterion (6) reveals that the number of experiments required to certify a system based on statistical sampling alone is of the order of $\frac{1}{2}\epsilon^{-2}\log\frac{1}{\epsilon'}$. The number-of-tests requirement is shown in Table I as a function of the probability-of-failure tolerance ϵ with $\epsilon' = \epsilon$. It is clear from this table that the number of tests required for the purely empirical certification of a system can be prohibitively expensive if the tests are costly and the required probability of failure is low.

It is natural to require that the confidence level on the estimation of the probability of failure is at least of the order of the required maximum probability of failure. That is why $\epsilon' = \epsilon$ is a natural chose. For $\epsilon' > \epsilon$ the uncertainty on

the estimation of the probability of failure would be larger than the probability of failure itself, which translates into the fact that the probability of failure may with probability ϵ' be above the certification threshold ϵ.

When empirical certification is not an option, the question that naturally arises is how models can be employed to reduce the number of tests required for certification. Thus, the goal of *model-based certification* is to achieve certification with a minimum of testing. In order to frame this question in mathematical terms, suppose that the behavior of the system is exactly described by an unknown function $Y = G(X, Z)$ of random input variables (X, Z). For definiteness, we shall assume that X and Z are independent and range over known intervals, and that no additional statistical information about the input parameters is available. These assumptions can be readily relaxed, cf. Appendix A, but such extensions will not be pursued here in the interest of simplicity. Suppose in addition that the behavior of the system is modeled by means of a function $Y = F(X)$. Thus X collects the input variables that are accounted for by the model, whereas Z collects those variables that are unaccounted for, or *unknown unknowns*. Evidently, if an exact model were available no testing would be required to achieve certification. In general, models cannot be expected to be exact and the degree of predictiveness of the models needs to be carefully assessed, a process known as *verification and validation*.

The manner in which concentration-of-measure supplies probability-of-failure upper bounds with the aid of validated models is summarized in this section through a sequence of representative scenarios of increasing complexity.

2.1 Scenario 1: Exact model, single performance measure whose mean is known

We begin by assuming that the mean performance $\mathbb{E}[Y]$ is known exactly and the model is perfect, i. e., there exists a random vector $X : \Omega \to \chi_1 \times \cdots \times \chi_M$ and a known function $F : \chi_1 \times \cdots \times \chi_M \to E_1 \times \cdots \times E_N$ such that the relation $Y = F(X)$ describes the system exactly. These assumptions represent ideal conditions in which all uncertainty regarding the response of the system is *aleatoric uncertainty*, i. e., stems from the stochastic variability of the system, and there is no *epistemic uncertainty*, i. e., the behavior of the system is known exactly, including its mean response. We also begin by considering the case in which certification depends on a single performance measure, i. e., $N = 1$. Under these assumptions, the admissible set is of the form $A = [a, \infty)$, where a is the minimum threshold for safe system operation. Then, if F is integrable (and that is the only assumption on F) and

the input parameters are independent McDiarmid's inequality states that

$$\mathbb{P}\big[F(X) - \mathbb{E}[F(X)] \leq -r\big] \leq \exp\big(-2\frac{r^2}{D_F^2}\big) \tag{7}$$

where

$$D_F^2 := \sum_{k=1}^{M} \sup_{(x_1,\ldots,x_{k-1},x_{k+1},\ldots x_M)\in\chi_1\times\ldots\chi_{k-1}\times\chi_{k+1}\times\cdots\times\chi_M} \sup_{(A_k,B_k)\in\chi_k^2} |F(x_1,\ldots,A_k,\ldots x_M) - F(x_1,\ldots,B_k,\ldots x_M)|^2 \tag{8}$$

is the *verification diameter* of the system. Bound (7) can be re-written in the form

$$\mathbb{P}[Y \in A^c] \leq \exp\big(-2\frac{(\mathbb{E}[Y]-a)_+^2}{D_F^2}\big) \tag{9}$$

where we write $x_+ := \max(0, x)$, whence it follows that the inequality

$$\frac{(\mathbb{E}[Y]-a)_+}{D_F} \geq \sqrt{\log\sqrt{\frac{1}{\epsilon}}} \tag{10}$$

supplies a conservative certification criterion.

Comparison of (10) and (1) affords the identification:

$$M = (\mathbb{E}[Y]-a)_+ \tag{11a}$$

$$U = D_F \tag{11b}$$

Thus, in the absence of epistemic uncertainty, i. e., for systems for which an exact model is available and whose mean performance is exactly known, *the margin M is the difference between the mean performance and the threshold*, or zero if this difference is negative, and *the uncertainty U equals the verification diameter of the system*. With these identifications, the certification criterion can be expressed in the form

$$\mathrm{CF} = \frac{M}{U} \geq \sqrt{\log\sqrt{\frac{1}{\epsilon}}} \tag{12}$$

This inequality establishes a clear correspondence between the probability-of-failure tolerance ϵ and the confidence factor CF. This correspondence is shown in tabular form in Table II. Thus, concentration-of-measure inequalities supply precise definitions of margin measures, uncertainty measures and minimum confidence factors that guarantee the safe operation of the system to within a pre-specified probability-of-failure tolerance.

failure tolerance (ϵ)	1	10^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}
confidence factor (CF)	0	1.07298	1.51743	1.85846	2.14597	2.39926	2.62826

Table II. Minimum confidence factor CF required to stay within a pre-specified probability-of-failure tolerance ϵ.

Several additional aspects of the certification method just described are noteworthy. Firstly, the only information about the model that is required for certification is the mean performance and the diameter of the system. In particular, the response function $F(X)$ need not be interpolated or otherwise represented, and can effectively be treated as a black box. Secondly, only ranges of input parameters, and not their detailed probability distribution functions, need be known for certification. Thirdly, the present scenario provides an extreme example of how the availability of a high-fidelity model helps to reduce the number of tests required for certification: when an exact model is available, the need for experimental testing is eliminated altogether.

It is also interesting to note that the verification diameter (8), which provides a rigorous measure of the aleatoric uncertainty in the response of the system, represents the largest deviation in system performance that is recorded when each input parameter is allowed to vary in turn between pairs of values spanning its entire range. Evidently, the computation of the verification diameter of a system entails an optimization over input-parameter space seeking to identify those large deviations in the input parameters that result in the largest deviations in the output parameters. It bears emphasis that consideration of large deviations is essential and that, in particular, linearized sensitivity analysis is not sufficient for rigorous certification in general. Specific optimization algorithms for the computation of verification diameters are discussed in Section 3.

It should be noted that we use the term *verification* in a somewhat expanded sense relative to other conventional uses of the term (cf., e. g., [4], [5]). Thus, a common use of the term verification is to signify the process of assessing how well a numerical model approximates the underlying physical laws governing the system, often expressed as a system of partial differential equations. However, in the present context the term verification naturally refers to the process of quantifying all aleatoric uncertainties, whether arising from numerical errors, from the statistical variability of the input parameters, the intrinsic stochasticity of the model, or from other sources. Thus, verification aims to quantify how precisely a model can predict the response of the system. A distinguishing characteristic of verification is that it is achieved solely by exercising the model and without reference to experimental data. The concentration-of-measure framework renders verification, often a conceptually appealing but ambiguous and imprecise term, in precise

quantitative terms. Thus, as already noted, verification is rigorously quantified by the verification diameter, in the sense that once the verification diameter of the system is known, the system can be certified rigorously and conservatively.

For completeness and in order to illustrate the assumptions required for (7) to hold, we include a simple version of McDiarmid's inequality [10] (where the constant 2 in the exponential has been replaced by $\frac{1}{2}$), also known as bounded differences inequality, and a sketch of its proof. The sharpening of inequality (14) to (7) is technically involved and we refer to [10] for a complete account. Independence is not a necessary condition for McDiarmid's inequality, we refer to [11] an extension of that inequality to centering sequences (for instance to martingales).

Theorem 2.1 *Assume that the random variables $X_1, \dots, X_M$ are independent. Then*

$$\mathbb{P}[F(X) - \mathbb{E}[F] \geq r] \leq \exp(-\frac{1}{2}\frac{r^2}{D_F^2}) \tag{13}$$

and

$$\mathbb{P}[F(X) - \mathbb{E}[F] \leq -r] \leq \exp(-\frac{1}{2}\frac{r^2}{D_F^2}) \tag{14}$$

Proof. Here we adapt the proof of corollary 1.17 in [6]. By Chebyshev's inequality we have

$$\mathbb{P}[F(X) - \mathbb{E}[F] \geq r] \leq e^{-\lambda r}\mathbb{E}\Big[\exp\big(\lambda(F - \mathbb{E}[F])\big)\Big] \tag{15}$$

for all $\lambda \geq 0$. In addition,

$$\mathbb{E}\Big[\exp\big(\lambda(F-\mathbb{E}[F])\big)\Big] = \mathbb{E}\Big[\mathbb{E}\big[\exp(\lambda(F-\mathbb{E}[F|\mathcal{F}_{n-1}]))|\mathcal{F}_{n-1}\big]\exp(\lambda(\mathbb{E}[F|\mathcal{F}_{n-1}]-\mathbb{E}[F]))\Big] \tag{16}$$

where $\mathcal{F}_k$ denotes the σ-algebra generated by $X_1, \dots, X_k$. We recall that, for every bounded function f of zero mean with respect to a measure ν and all $\lambda \geq 0$, Jensen's inequality gives

$$\int e^{\lambda f} d\nu \leq \int\int e^{\lambda(f(x)-f(y))} d\nu(x) d\nu(y) \leq \sum_{i=0}^{\infty} \frac{(D_f\lambda)^{2i}}{(2i)} \leq \exp(D_f^2\lambda^2/2) \tag{17}$$

with $D_f := \sup_{x,y}|f(x) - f(y)|$. Applying inequality (17) to the integration with respect to the law of X_n we obtain

$$E\big[\exp(\lambda(F - \mathbb{E}[F|\mathcal{F}_{n-1}]))|\mathcal{F}_{n-1}\big] \leq \exp(D_n^2\lambda^2/2) \tag{18}$$

where

$$D_n := \sup_{x_1,\ldots,x_{n-1},x,y} |F(x_1,\ldots,x_{n-1},x) - F(x_1,\ldots,x_{n-1},y)| \tag{19}$$

It follows that

$$\mathbb{E}\Big[\exp\big(\lambda(F-\mathbb{E}[F])\big)\Big] = \exp(D_n^2\lambda^2/2)\mathbb{E}\Big[\exp(\lambda(\mathbb{E}[F|\mathcal{F}_{n-1}]-\mathbb{E}[F]))\Big] \tag{20}$$

and, by induction,

$$\mathbb{E}\Big[\exp\big(\lambda(F-\mathbb{E}[F])\big)\Big] = \exp(D_F^2\lambda^2/2) \tag{21}$$

Combining this inequality with Eq. (15) and taking $\lambda = r/D_F^2$ finally gives

$$\mathbb{P}[F(X)-\mathbb{E}[F]\geq r] \leq \exp(-\frac{1}{2}\frac{r^2}{D_F^2}) \tag{22}$$

Inequality 14 is obtained by replacing F by $-F$. □

We additionally observe that Hoeffding's inequality [9]

$$\mathbb{P}[\frac{1}{M}\sum_{i=1}^{M} X_i - \frac{1}{M}\sum_{i=1}^{M}\mathbb{E}[X_i]\geq r] \leq \exp(-2M\frac{r^2}{(\sum_{i=1}^{M}(b_i-a_i)^2/M)}) \tag{23}$$

follows from McDiarmid's inequality as a special case when the spaces E_i are equal to intervals (a_i, b_i) and

$$F(X) = \frac{1}{M}\sum_{i=1}^{M} X_i \tag{24}$$

2.2 Scenario 2: Exact model, multiple performance measures whose mean is known

The preceding framework can be extended to the case in which certification depends on more than one performance measure. To this end, suppose that A is an arbitrary subset of $\mathbb{R}^N$, i. e., $N \geq 1$ and $E_i = \mathbb{R}$, $i = 1,\ldots,N$, and that the mean performance $\mathbb{E}[Y]$ is known and belongs to the interior of A. Then, the following concentration-of-measure inequality is deduced from McDiarmid's inequality

$$\mathbb{P}[Y\in A^c] \leq \inf_{s,r\in\mathbb{R}^N\,:\,\mathbb{E}[Y]+\prod_{i=1}^{N}(-s_i,r_i)\subset A} \sum_{i=1}^{N}\left[\exp\left(-2\frac{r_i^2}{D_{F_i}^2}\right)+\exp\left(-2\frac{s_i^2}{D_{F_i}^2}\right)\right] \tag{25}$$

where D_{F_i} is the verification diameter of the response function F_i and the infimum is taken over all hyperrectangles contained in A. This inequality follows simply

from McDiarmid's inequality by observing that, for all $s, r \in \mathbb{R}^N$ ($s := (s_1, \ldots, s_N)$) such that $\mathbb{E}[Y] + \prod_{i=1}^N (-s_i, r_i) \subset A$, one has

$$A^c \subset \cup_{i=1}^N \{Y_i - \mathbb{E}[Y_i] \geq r_i\} \cup \{Y_i - \mathbb{E}[Y_i] \leq -s_i\}, \tag{26}$$

whence it follows that

$$\mathbb{P}[Y \in A^c] \leq \sum_{i=1}^N \big(\mathbb{P}[Y_i - \mathbb{E}[Y_i] \geq r_i] + \mathbb{P}[Y_i - \mathbb{E}[Y_i] \leq -s_i]\big). \tag{27}$$

As in the preceding scenario, the inequality

$$\inf_{s,r \in \mathbb{R}^N \,:\, \mathbb{E}[Y] + \prod_{i=1}^N (-s_i, r_i) \subset A} \sum_{i=1}^N \left[\exp\left(-2\frac{r_i^2}{D_{F_i}^2}\right) + \exp\left(-2\frac{s_i^2}{D_{F_i}^2}\right)\right] \leq \epsilon \tag{28}$$

now supplies a conservative certification criterion. It follows that in the case of multiple performance measures certification can be achieved by the computation of the verification diameters of each of the components of the response function.

Suppose, for example, that $A = \prod_{i=1}^N (a_i, +\infty)$, where a_i is the threshold of the ith performance measure, and suppose that $a_i \leq \mathbb{E}[Y_i]$, $i = 1, \ldots, N$. Then, the certification inequality (28) reduces to

$$\sum_{i=1}^N \exp\left(-2\frac{(\mathbb{E}[Y_i] - a_i)_+^2}{D_{F_i}^2}\right) \leq \epsilon \tag{29}$$

It should also be noted that Equation (29) is based on the worst case scenario that these events do not intersect and therefore the sum of their respective probabilities is an upper bound for the union. In practical applications, there may be an overlap between the individual failure events that constitutes this bound. This point will be addressed in a sequel work.

By analogy to the case of a single performance measure we can now introduce the margins and uncertainty measures

$$M_i = (\mathbb{E}[Y_i] - a_i)_+ \tag{30a}$$

$$U_i = D_{F_i} \tag{30b}$$

for each performance measure in turn. Then, (29) can be rewritten in the form

$$\mathrm{CF} = \sqrt{\log\left(\frac{1}{\sqrt{\sum_{i=1}^N \exp\left(-2(\mathrm{CF}_i)^2\right)}}\right)} \geq \sqrt{\log\sqrt{\frac{1}{\epsilon}}} \tag{31}$$

where we write

$$\mathrm{CF}_i = \frac{M_i}{U_i} \tag{32}$$

CF_i	1.0	1.2	1.4	1.6	1.8	2.0
1.0	0.808348	0.909127	0.965192	0.989139	0.997179	0.999381
1.2	0.909127	1.045670	1.135196	1.178736	1.194371	1.198758
1.4	0.965192	1.135196	1.270207	1.352168	1.386639	1.397003
1.6	0.989139	1.178736	1.352168	1.487759	1.563896	1.591443
1.8	0.997179	1.194371	1.386639	1.563896	1.701007	1.772316
2.0	0.999381	1.198758	1.397003	1.591443	1.772316	1.911394

Table III. Aggregate confidence factor for a system characterized by two performance measures with confidence factors CF_1 (rows) and CF_2 (columns).

Evidently, the certification criterion (31) reduces to (12) in the case of a single performance measure. It is interesting to note that, in the case of multiple performance measures, the confidence factor CF follows as an aggregate of the confidence factors CF_i of each of the performance measures according to the composition rule (31). However, it should be carefully noted that neither margins nor uncertainties can be aggregated independently of each other to define an overall margin and an overall uncertainty of the system. The confidence-factor aggregation relation is shown in Table III for a system characterized by two performance measures. As expected, the aggregate confidence factor is smaller than each of the confidence factors corresponding to the individual performance measures. Thus, lack of confidence in individual performance measures compounds and the overall confidence in the system takes a hit with the addition of every new performance measure. However, because the individual confidence factors enter the aggregation relation (31) through exponentials, it follows that the aggregate confidence factor–and the certification process itself–is dominated by those performance measures having the smallest individual confidence factors. Conversely, performance measures having large confidence factors have negligible effect on the overall confidence factor of the system and can be safely removed from consideration in the certification process.

2.3 Scenario 3: Exact model, single performance measure whose mean is unknown

In the foregoing we have assumed that the mean performance $\mathbb{E}[Y]$ of the system is known *a priori*. However, in most situations of practical interest such information is not available and the mean performance must be estimated instead. Suppose that, to this end, we perform m evaluations of the model $F(X)$ based on an unbiased sampling of the input parameters, resulting in predicted performances $Y^1, Y^2, \ldots, Y^m$. We then define the *estimated mean performance* corresponding

to these calculations as

$$\langle Y \rangle = \frac{1}{m} \sum_{i=1}^{m} Y^i \tag{33}$$

Start by additionally assuming that there is one single performance measure, $N = 1$, and that the safe operation of the system requires that $Y \geq a$ for some threshold a, i. e., $A = [a, \infty)$. The probability $\mathbb{P}[Y \in A^c]$ can now only be determined to within confidence intervals reflecting the randomness of the estimated mean $\langle Y \rangle$. Under these conditions, we obtain the following inequality

$$\mathbb{P}\left[\mathbb{P}[Y \in A^c] \geq \exp\left(-2\frac{[\langle Y \rangle - a - \alpha]_+^2}{D_F^2}\right)\right] \leq \epsilon'. \tag{34}$$

where ϵ' is a pre-specified *estimation tolerance* and

$$\alpha = D_F m^{-\frac{1}{2}} (-\log \epsilon')^{\frac{1}{2}}. \tag{35}$$

Inequality (34) follows simply from an application of McDiarmid's inequality to $\langle Y \rangle$, with the result

$$\mathbb{P}[\mathbb{E}[Y] - \langle Y \rangle \leq -\alpha] \leq \epsilon' \tag{36}$$

whence it follows that, with $\mathbb{P}$ probability $1 - \epsilon'$,

$$A^c \subset \{Y - \mathbb{E}[Y] \leq a + \alpha - \langle Y \rangle\}. \tag{37}$$

Next we observe that (34) simply states that, with probability $1 - \epsilon'$,

$$\mathbb{P}[Y \in A^c] \leq \exp\left(-2\frac{[\langle Y \rangle - a - \alpha]_+^2}{D_F^2}\right) \tag{38}$$

and the certification criterion (3) now becomes

$$\frac{[\langle Y \rangle - a - \alpha]_+}{D_F} \geq \sqrt{\log \sqrt{\frac{1}{\epsilon}}}\,. \tag{39}$$

This certification criterion is again of the form (12) with margin and uncertainty measure

$$M = [\langle Y \rangle - a - \alpha]_+ \tag{40a}$$

$$U = D_F \tag{40b}$$

Comparison of (11) and (40) shows that the estimation of the mean performance of the system effectively reduces the margin in the amount α. Evidently, this margin hit can be reduced to an arbitrarily small value by carrying out a sufficiently large number of model evaluations. The certification criterion (40) again shows that, as in the case of known mean performance, in the absence of epistemic uncertainty, certification can be rigorously achieved from the sole knowledge of the verification diameter of the system and an estimate of its mean performance.

2.4 Scenario 4: Exact model, multiple performance measures whose means are unknown

For completeness, we proceed to record the extension of the preceding case to multiple performance measures, $N \geq 1$, and arbitrary A in $\mathbb{R}^N$. In this case, with probability $1-\epsilon'$,

$$\mathbb{P}[Y \in A^c] \leq \inf_{s,r\in\mathbb{R}^N\,:\,\mathbb{E}[Y]+\prod_{i=1}^N(-s_i-\alpha_i,r_i+\alpha_i)\subset A} \sum_{i=1}^N \Big(\exp\big(-2\frac{r_i^2}{D_{F_i}^2}\big) + \exp\big(-2\frac{s_i^2}{D_{F_i}^2}\big)\Big) \tag{41}$$

where $\alpha \in \mathbb{R}^N$ and

$$\alpha_i = D_{F_i}\big(\frac{\log\frac{2N}{\epsilon'}}{2m}\big)^{\frac{1}{2}} \tag{42}$$

Inequality (41) again follows from McDiarmid's inequality by observing that

$$\mathbb{P}[|\mathbb{E}[Y^i] - \langle Y\rangle^i| \geq \alpha_i] \leq \frac{\epsilon'}{N}. \tag{43}$$

Hence, $|\mathbb{E}[Y^i] - \langle Y\rangle^i| \leq \alpha_i$ for all i with probability $1-\epsilon'$, with the result that for all $s, r \in \mathbb{R}^N$ such that $\langle Y\rangle + \prod_{i=1}^N(-s_i-\alpha_i, r_i+\alpha_i) \subset A$,

$$\{Y \in A^c\} \subset \cup_{i=1}^N \{Y_i - \mathbb{E}[Y_i] \geq s_i\} \cup \{\mathbb{E}[Y_i] - Y_i \geq r_i\}. \tag{44}$$

with probability $1-\epsilon'$, and the certification criterion finally becomes

$$\inf_{s,r\in\mathbb{R}^N\,:\,\langle Y\rangle+\prod_{i=1}^N(-s_i-\alpha_i,r_i+\alpha_i)\subset A} \sum_{i=1}^N \Big(\exp\big(-2\frac{r_i^2}{D_{F_i}^2}\big) + \exp\big(-2\frac{s_i^2}{D_{F_i}^2}\big)\Big) \leq \epsilon \tag{45}$$

For example, suppose, as in scenario 2, that $A = \prod_{i=1}^N (a_i, +\infty)$, where a_i is the threshold of the ith performance measure, and suppose that $a_i + \alpha_i \leq \langle Y_i\rangle$, $i = 1, \ldots, N$. Then, the certification inequality (45) reduces to

$$\sum_{i=1}^N \exp\left(-2\frac{(\langle Y_i\rangle - a_i - \alpha_i)_+^2}{D_{F_i}^2}\right) \leq \epsilon \tag{46}$$

where now

$$\alpha_i = D_{F_i}\big(\frac{\log\frac{N}{\epsilon'}}{2m}\big)^{\frac{1}{2}} \tag{47}$$

in view of the one-side form of the admissible intervals. By analogy to the case of a single performance measure we can now introduce the margins and uncertainty measures

$$M_i = (\langle Y_i\rangle - a_i - \alpha_i)_+ \tag{48a}$$

$$U_i = D_{F_i} \tag{48b}$$

for each performance measure in turn. Then, (46) can be rewritten in the form (31) with the individual performance measure confidence factor defined as in (32). As in the case of a single performance measure, we observe that the need to estimate the mean performance has the effect of reducing the individual performance measure margins in the amounts α_i. These are controllable margin hits that can be reduced to any desired extent by carrying out a sufficiently large number of model evaluations. Again, in the absence of epistemic uncertainty certification can be rigorously achieved from the sole knowledge of the verification diameters of the individual performance measures and estimates of their mean performance.

2.5 Scenario 5: Inexact model

In practice F stands for a numerical or analytical model of a physical system whose output is the random vector Y. The model accounts for some of the input parameters X that determine the performance of the system. If the model were perfect then, for all X, $F(X)$ would exactly equal the outcome Y of an experiment performed with an identical set of input parameters. In general, Y and $F(X)$ are not equal owing to:

i) imperfections in the model,

ii) the existence of additional unknown random input parameters, or *unknown unknowns*, not accounted for in the model.

These two sources of error and epistemic uncertainty can be analyzed by supposing that the exact response of the physical system is governed by a function $G(X, Z)$, generally unknown, of the random variables X that are accounted for by the model and additional unknown unknowns Z. Even if no unknown unknowns exist and the model accounts for all input parameters of the system, in general $G(X, Z) \neq F(X)$ owing to the limited fidelity of the model $F(X)$. Evidently, these sources of error and epistemic uncertainty, namely, the existence of unknown unknowns and the limited fidelity of the model, must be carefully assessed as part of the certification of the system.

To this end, begin by considering the case of a single performance measure, $N = 1$, and by noting that

$$\{Y \leq a\} \subset \{F(X) \leq a + h\} \cup \{G(X, Z) - F(X) \leq -h\}\,, \tag{49}$$

where $G - F$ may be regarded as a *modeling-error function* and h is an arbitrary number, leading to the estimate

$$\mathbb{P}[Y \leq a] \leq \mathbb{P}[F(X) \leq a + h] + \mathbb{P}[G(X, Z) - F(X) \leq -h] \tag{50}$$

and to the conservative certification criterion

$$\mathbb{P}[F(X) \leq a + h] + \mathbb{P}[G(X, Z) - F(X) \leq -h] \leq \epsilon \tag{51}$$

Ideally h should be chosen to minimize the sum of the probabilities on the left hand side of Eq. (51).

We have seen in previous sections how to obtain a bound for $\mathbb{P}[F(X) \leq a+h]$, i. e., for the probability of failure predicted by the model (with an additional margin h). Certification now additionally requires a bound on $\mathbb{P}[G(X, Z) - F(X) \leq -h]$, i. e., the probability that the predicted and measured performance differ significantly. This probability measures the deleterious effect on predictiveness of all model imperfections, whether resulting from the limited fidelity of the model or from unknown unknowns, and may therefore be regarded as a measure of *epistemic uncertainty*. As is evident from (51), the epistemic uncertainty has the effect of decreasing the effective probability-of-failure tolerance. In particular, certification is not possible if the model is not sufficiently faithful, i. e., if $\mathbb{P}[G(X, Z) - F(X) \leq -h] \geq \epsilon$.

The epistemic uncertainty can again be rigorously bounded by means of concentration-of-measure inequalities. Thus, a direct application of McDiarmid's inequality to $G - F$ gives the concentration-of-measure bound

$$\mathbb{P}[G(X, Z) - F(X) \leq -h] \leq \exp\big(-2\frac{(\mathbb{E}[G - F] + h)_+^2}{D_{G-F}^2}\big) \tag{52}$$

where the *validation diameter*

$$\begin{aligned} D_{G-F}^2 := \sum_{k=1}^{M} & \sup_{(x_1,\ldots,x_{k-1},x_{k+1},\ldots x_M)\in E_1\times\ldots E_{k-1}\times E_{k+1}\times\cdots\times E_M} \sup_{z\in E_{M+1}} \sup_{(A_k,B_k)\in E_k^2} \\ & |F(x_1,\ldots,x_{k-1},A_k,x_{k+1},\ldots x_M) - G(x_1,\ldots,x_{k-1},A_k,x_{k+1},\ldots x_M,z) \\ & - F(x_1,\ldots,x_{k-1},B_k,x_{k+1},\ldots x_M) + G(x_1,\ldots,x_{k-1},B_k,x_{k+1},\ldots x_M,z)|^2 \\ & + \sup_{(x,z)\in E_1\times\cdots\times E_{M+1}, z'\in E_{M+1}} |G(x,z) - G(x,z')|^2 \end{aligned} \tag{53}$$

supplies a measure of the *epistemic uncertainty* of the system. In particular, the validation diameter measures the extent to which the predictions of the model deviate from observation.

In practice $\mathbb{E}[G - F]$ and $\mathbb{E}[F]$ are not known and must instead be estimated. let $\langle G - F\rangle$ and $\langle F\rangle$ be the empirical means of $G - F$ and F, respectively, estimated from m nominally identical model evaluations and experiments. Take $\alpha_F := D_F m^{-\frac{1}{2}}(-\log\epsilon')^{\frac{1}{2}}$ and $\alpha_{G-F} := D_{G-F} m^{-\frac{1}{2}}(-\log\epsilon')^{\frac{1}{2}}$. Then we have seen

that, with probability $1-2\epsilon'$ for all $h \in [0,a]$,

$$\mathbb{P}[F(X) \leq a+h] \leq \exp\big(-2\frac{(\langle F\rangle - a - h - \alpha_F)_+^2}{D_F^2}\big) \tag{54}$$

and

$$\mathbb{P}[G(X,Z) - F(X) \leq -h] \leq \exp\big(-2\frac{(\langle G-F\rangle + h - \alpha_{G-F})_+^2}{D_{G-F}^2}\big) \tag{55}$$

It therefore follows that the inequality

$$\inf_h \exp\big(-2\frac{(\langle F\rangle - a - h - \alpha_F])_+^2}{D_F^2}\big) + \exp\big(-2\frac{(\langle G-F\rangle + h - \alpha_{G-F})_+^2}{D_{G-F}^2}\big) \leq \epsilon \tag{56}$$

supplies a conservative certification criterion. A near optimal choice for h is given by matching the expressions in the exponentials, with the result

$$h = \frac{(\langle F\rangle - a - \alpha_F)D_{G-F} + (\alpha_{G-F} - \langle G-F\rangle)D_F}{D_F + D_{G-F}} \tag{57}$$

whence the certification criterion becomes

$$CF = \frac{M}{U} \geq \sqrt{\log\sqrt{\frac{2}{\epsilon}}} \tag{58}$$

with

$$M = \big(\langle F\rangle - a - \alpha_F - \alpha_{G-F} + \langle G-F\rangle\big)_+ \tag{59a}$$

$$U = D_F + D_{G-F} \equiv U_A + U_E \tag{59b}$$

It is interesting to note that the total uncertainty U is indeed the sum of an aleatoric uncertainty U_A, measured by the verification diameter D_F, and an epistemic uncertainty U_E, measured by the validation diameter D_{G-F}. Generalizations to the cases of multiple performance measures, $N > 1$, estimated mean performance and arbitrary admissible set follow along similar lines as those presented in the preceding scenarios.

We also observe that the effect of epistemic uncertainty, whether due to lack of fidelity of the model or the existence of unknown unknowns, is to decrease the margin M by $\alpha_{G-F} + \langle G-F\rangle$ and increase the total uncertainty U by D_{G-F}. It should also be carefully noted that the determination of the validation diameter D_{F-G} requires the simultaneous and coordinated execution of the model F and experiments G for equal known parameters X, in order to assess the fidelity of the model, and the repetition of experiments for equal known parameters X, in

order to assess the effect of the unknown unknowns. Since experiments are often costly and time consuming, the value and practicality of model-based certification thus depends critically on the ability to quantify epistemic uncertainties by means a sufficiently small number of experiments. However, for a sufficiently predictive model the function $G - F$ involved in the evaluation of validation diameter D_{G-F} can be expected to exhibit much less variation than either F or G, thus enabling the computation of D_{G-F} by means of a rapidly-converging iterative scheme requiring a small number of model evaluations and experiments.

Thus, the precise manner in which a predictive model cuts down on the number of experiments required for certification is by restricting the need of testing to the specific purpose of *validation*, presumably a much less testing-intensive task than purely empirical certification of the system regarded as a black box. Evidently, the requisite number of experiments depends critically on the quality of the model, as well as on the method of optimization used to determine the epistemic uncertainties. An extreme case is furnished by a perfect model that accounts for all input parameters, in which case the quantification of the total uncertainty requires no experiments. In general, the purpose and benefit of model-based certification is to reduce the number of experiments required for certification through the formulation of a sufficiently predictive model. Herein lies the promise, as well as the challenge, of model-based certification.

3 Numerical tests

In this section we apply, by way of demonstration, the concentration-of-measure approach to the problem of predicting the state of maximum compression imploding ring. The behavior of the ring in the high-energy regime is strongly nonlinear and dynamical and the calculation of performance measures requires the solution of an initial value problem in time. Because the system undergoes multiple bifurcations and generally traverses a rough energy landscape, the performance measures are expected to depend sensitively on initial conditions and on the parameters of the system. In addition, in the high-energy regime the shape of the ring at maximum compression is expected to be highly crumpled and irregular. Thus, while straightforward in its definition, the example of an imploding ring does pose a non-trivial and illuminating test of the theory.

That the certification criterion (Eq. (39)) is indeed conservative, i. e., that if Eq. (39) is satisfied then the probability-of-failure is indeed less than ϵ with probability $1 - \epsilon'$, has been rigorously proven mathematically. However, several practical questions remain to be ascertained. A first question concerns whether the bounds furnished by concentration-of-measure are tight enough to supply a practical means of certification. A second question concerns the means of calculation of

the aleatoric and epistemic uncertainties, as measured by their corresponding verification and validation diameters, including the relative efficiencies of optimization algorithms and the number of system evaluations and experimental tests required for the computation of the uncertainty measures. The imploding ring example presented in this section sheds useful light on these and other related questions.

3.1 Test case description

We consider a ring described by means of a bead model consisting of n point masses interacting through two- and three-body potentials. The ring starts from an equilibrium circular configuration and is imparted an inward radial velocity. The objective of the analysis is to characterize the state of the ring at the point of maximum compression.

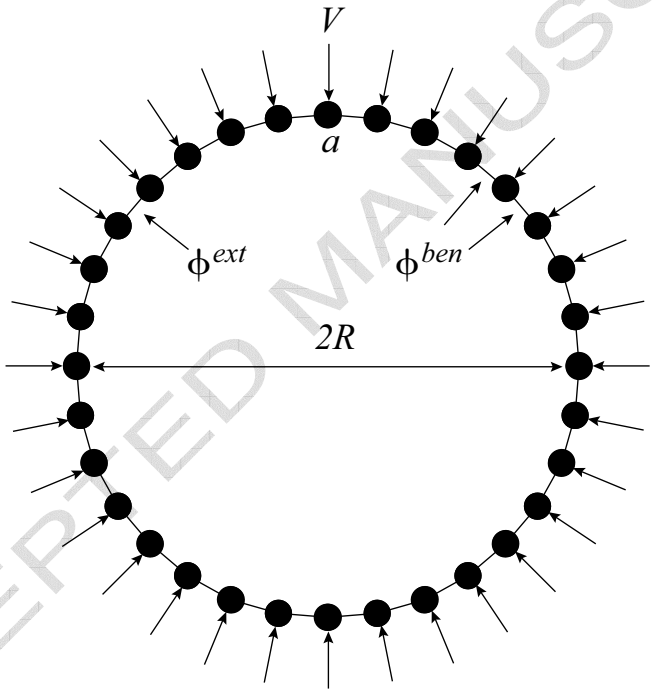


Figure 1: Schematic of ring implosion and explosion test.

The configuration of the ring at time t is characterized by the coordinates $x_a(t)$, $a = 1, \ldots, n$, of all the beads. The total energy is assumed to be of the form

$$E(x) = E^{\text{ext}}(x) + E^{\text{ben}}(x) = \sum_{\text{bonds}} \phi^{\text{ext}} + \sum_{\text{masses}} \phi^{\text{ben}} \tag{60}$$

where for notational convenience we collect in the array $x = \{x_a, a = 1, \ldots, n\}$ the coordinates of all the beads, ϕ^{ext} is a two-body potential giving the extensional energy of a bond, ϕ^{ben} is a three-body potential giving the bending energy of a pair of bonds incident on a bead, and E^{ext} and E^{ben} are the total extensional and bending energies of the ring respectively. In calculations, we adopt the simple two-body potential

$$\phi^{\text{ext}} = \frac{C}{2} \log^2 \left(\frac{r}{a_0} \right) \tag{61}$$

where r is the length of the bond and a_0 and C are the equilibrium length and the extensional stiffness of the bond, respectively, and the simple three-body potential

$$\phi^{\text{ben}} = \frac{D}{2} \left(\frac{r_1 r_2 \cos\alpha - a_1 a_2 \cos\alpha_0}{r_1 r_2 \cos\alpha + a_1 a_2 \cos\alpha_0} \right)^2 \tag{62}$$

where r_1 and r_2 are the lengths of the two bonds incident on a bead, a_1 and a_2 are their respective equilibrium lengths, D is the bond-pair bending stiffness, α is the angle subtended by the bonds and α_0 is the corresponding equilibrium angle. The two-body force diverges strongly as $r \to 0$, which prevents bond collapse. By contrast, the two-body force goes through a maximum at $r = ea_0$, and the three-body force resulting from the extension of the bonds at constant angle likewise goes through a maximum at $r = 2a_0$, which allows for bond breaking in tension. The three-body potential diverges when $\alpha \to \pi - \alpha_0$, with bonds otherwise unstretched, causing the bond angles to remain in the interval $(-\pi + \alpha_0, \pi - \alpha_0)$. The three-body potential is otherwise compliant about the equilibrium angle α_0, which allows for the development of fine waviness and renders the behavior of the system difficult to predict.

The resulting equations of motion of the ring are

$$M\ddot{x}(t) + f(x(t)) = 0 \tag{63}$$

where we use a superposed dot to denote differentiation with respect to time, the diagonal matrix M collects the masses of the beads, and

$$f(x) = \frac{\partial E}{\partial x}(x) \tag{64}$$

is the interaction force array. In order to approximate the trajectory of the ring we discretize the equations of motion, Eq. (63), in time by means of the standard explicit Newmark algorithm (cf, e. g., [12])

$$x_{n+1} = x_n + \Delta t v_n + \frac{1}{2}\Delta t^2 a_n \tag{65a}$$

$$v_{n+1} = v_n + \frac{\Delta t}{2}(a_n + a_{n+1}) \tag{65b}$$

$$a_{n+1} = -M^{-1} f(x_{n+1}) \tag{65c}$$

where Δt is the increment of time, and (x_n, v_n, a_n) denote the coordinates, velocities and accelerations of the ring at time $t_n = n\Delta t$. Eqs. (65) are *explicit* and can be used to advance the solution in time from the initial conditions. We specifically assume that the ring is initially at rest and in an equilibrium configuration in the form of a circle of radius R. In addition, we assume a initial radial velocity profile of the sinusoidal form

$$V(\theta) = V_0 + \Delta V \sin(k\theta) \tag{66}$$

where θ is the polar angle around the ring, V_0 represents the mean initial radial velocity, ΔV the amplitude of the fluctuation, and k is the wave number of the fluctuation.

The initial time step Δt is chosen to be of the order of one tenth of the linearly stable time step for explicit integration. In calculations we make use of a Richardson extrapolation time-step selection scheme to ensure that the total energy remains ostensibly constant. Specifically, the state of the ring is advanced twice from time t_n to t_{n+1} by: a single update with the current time step Δt_n; and two consecutive updates with time step $\Delta t_n/2$. If the energy error estimate at time t_{n+1} obtained by Richardson extrapolation is within a prescribed tolerance we set $\Delta t_{n+1} = \Delta t_n$ and proceed. In addition, the Richardson extrapolation of the two trajectories is used as the state of the ring at t_{n+1}. If the energy error estimate exceeds the prescribed tolerance the entire process is repeated with Δt_n halved. The iteration stops if the time step decreases below a minimum value, in which case the calculation is discarded.

A typical trajectory of the ring is shown in Fig. 2. As may be observed in the figure, the ring buckles early on and subsequently undergoes extensive crumpling, with the amplitude of the resulting crumpling increasing monotonically up to the point of maximum compression.

Suppose that the objective of the analysis is to predict the bending energy E^{ben} at the point of maximum compression of the trajectory. A natural measure of the extent of compression of the ring is supplied by the extensional energy E^{ext}. Likewise, the bending energy E^{ben} provides a natural measure of the degree of crumpling of the ring in compression. Therefore, an appropriate performance measure is given by the value of the bending energy E^{ben} of the ring at the time the extensional energy E^{ext} attains its first maximum. Thus, simulations stop when E^{ext} attains its first maximum in time, at which point the value of E^{ben} is recorded as the single performance measure Y of interest. We shall additionally assume that the proper operation of the system requires Y to be above a certain threshold a. Thus, the system fails when the bending energy achieved at the point of maximum compression falls short of the threshold value.

The calculations described earlier implicitly define a response function $Y = F(X)$ that returns the performance measure Y as a function of the parameters X

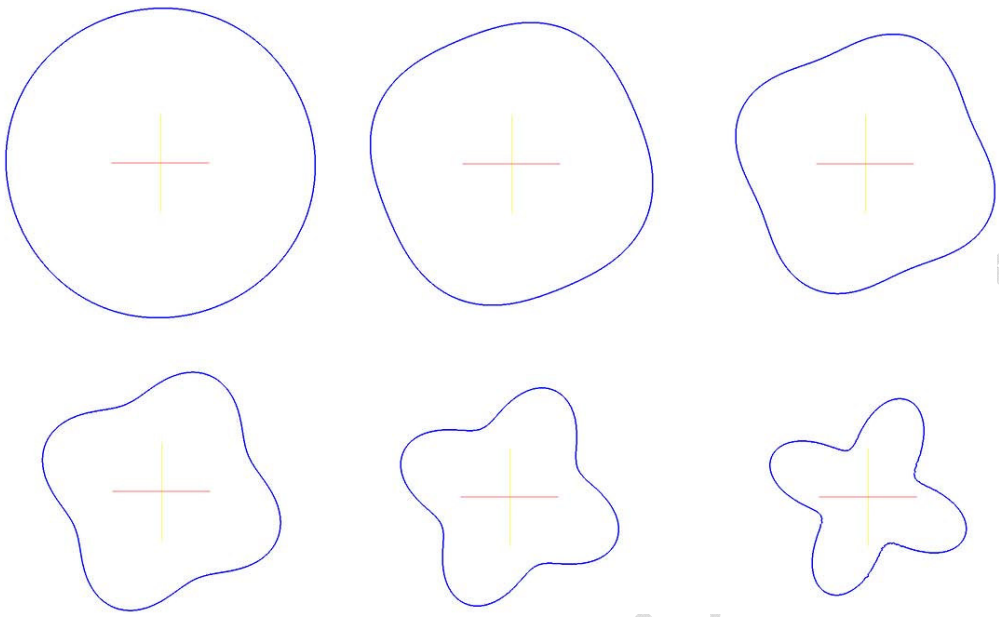


Figure 2: Ring implosion test. Crumpling of circular ring resulting from implosion (left to right, top to bottom).

of the system. These parameters include material constants, geometry, loading, initial conditions and numerical parameters such as the time step used for numerical integration. For definiteness, we shall regard all parameters to be certain, and therefore treat them as fixed constants, with the exception of the bending stiffnesses D of the bond-pairs in the ring, which we assume to be uncertain. Specifically, we divide the circumference of the ring into M segments of equal length. Each of the beads in a particular section $i = 1, \ldots, M$ is assigned a variable value X_i for the bending stiffness within a certain range. In order to investigate the effect of the range of the inputs on the probability-of-failure estimates we have considered the bending stiffness ranges $[0.95, 1.05]$, $[0.99, 1.01]$, $[0.995, 1.005]$, and $[0.999, 1.001]$. In addition, in order to examine the corresponding effect of the nonlinearity of the model we have considered the mean initial velocities $V_0 = 3$, 3×10^{-1}, 3×10^{-2}, and 3×10^{-3}. The remaining values of the parameters used in calculations are tabulated in Table IV.

Parameter	Value
Number of beads (n)	256
Extensional stiffness (C)	1.0
Individual bead mass density (m)	1.0
Ring radius (R)	0.25
Initial time step (Δt)	6.14×10^{-4}
Initial velocity fluctuation (ΔV)	$V_0/3.0$
Initial velocity wave number (k)	4
Minimum time step	1.0×10^{-6}
Richardson extrapolation energy tolerance	1.0×10^{-3}

Table IV. Values of fixed parameters used in imploding ring calculations.

3.2 Uncertainty quantification analysis

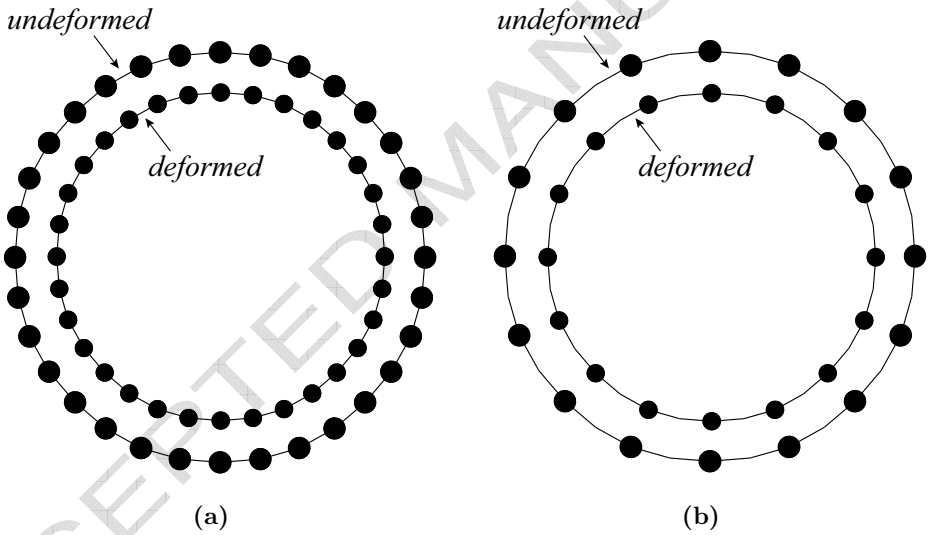


Figure 3: Ring configurations considered in the definition of the coarse-grained model. a) Exact model; b) Coarse-grained model.

The fundamental calculations to be carried out for purposes of uncertainty quantification are the calculation of the estimated mean performance $\langle Y \rangle$, Eq. (33), the verification diameter D_F of the response function, Eq. (8), and the validation diameter D_{G-F}, Eq. (53). In lieu of experimental data, G is taken as the response function of a higher resolution ring model discretized with 768 beads, with F then

representing a coarse ring discretized with 256 beads obtained by coarse-graining G. As in the case of actual experimental data, the evaluations of G are expensive, which places a premium on optimization methods that require the least number of evaluations of G. The extensional and bending energies of the coarse-grained model are taken to be of the same form (61) and (62), respectively, with effective axial and bending stiffness coefficients determined so that the energies of the two configurations shown in Fig. 3, corresponding to uniform compression or extension of the ring, are identical. In addition, we require the kinetic energies of the coarse and fine models to be identical for uniform velocity around the ring. Relations consistent with these constraints are

$$C_F = \sum C_G \,, \tag{67a}$$

$$D_F = \sum D_G \,, \tag{67b}$$

$$m_F = \sum m_G \,. \tag{67c}$$

Here C_F, D_F and m_F denote the axial stiffness, bending stiffness and bead in the coarse-grained model, C_G, D_G and m_G denote the axial stiffness, bending stiffness and bead mass in the exact model, and the sum extends to the beads in the fine model represented by each bead in the coarse model. We note that $\alpha = \alpha_0$, i. e., the bond angles remain unchanged, uniformly throughout the ring in the two configurations shown in Fig. 3, leading to the simple scaling (67). For large deviations from a perfect circular configuration, the coarse-grained model F deviates significantly from the exact model G, especially as a result of the strong nonlinear dependence of the bending energy on the bond-pair angle and of the inability of the coarse-grained geometry to resolve fine wrinkles in the deformed configuration of the exact model. It bears emphasis that the upscaling model just described is not proposed as an accurate method of coarse-graining but, contrariwise, it is intended as a rough approximation for purposes of ascertaining the uncertainties introduced by modeling assumptions.

In addition, in order to introduce unknown unknowns of the type that are likely to be encountered in practice, i. e., arising from fine unresolved length scales, we introduce nodal mass perturbations in fine model G that are subgrid relative to the coarse model F, i. e., preserve the aggregate masses (67c) of all beads in F. This constraint is satisfied by adding a mass Z to the beads of G explicitly accounted for in F, and subtracting masses $Z/2$ to the remaining beads in G. This distribution of mass over the fine model is indeed undetectable on the level of resolution of the coarse model. In this manner, the added mass Z becomes sole unknown unknown of the system. In calculations, the added mass Z is either zero or is varied bead-wise so that the variation of the total mass of the ring is in the range $[-0.1, 0.1]$.

The estimation of the mean performance $\langle Y \rangle$ can be carried out simply by means of Monte Carlo sampling. The computation of the diameters D_F and D_{G-F} requires an optimization over parameter space. Owing to the roughness of the energy function landscape of the ring and the lack of explicit derivatives of the response function, global optimization methods such as genetic algorithms and simulated annealing naturally suggest themselves in the computation of D_F. In calculations we employ the quasi-Newton method and genetic algorithms implemented in Sandia National Laboratories' DAKOTA Version 4.0 software package. By contrast, the computation of D_{G-F} requires costly experimental tests and global optimization algorithms are often not viable due to their slow convergence. The expectation, however, is that for a sufficiently high-fidelity model the function $G - F$ exhibits much less variation than either F or G, and that, therefore, the computation of D_{G-F} can be carried out by means of rapidly converging iterative schemes such as a quasi-Newton iteration.

3.3 Implosion tests

3.3.1 Perfect model

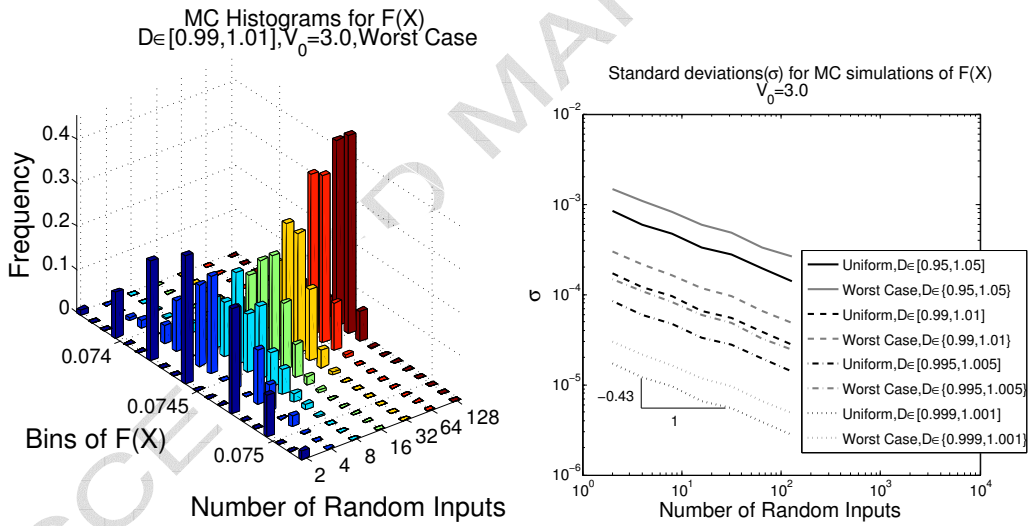


Figure 4: Ring implosion test. Monte Carlo calculations (10^3 samples per system) of the distribution of the performance measure as a function of the number of random variable inputs of the system (a) Histogram of the performance measure. (b) Standard deviation (σ) on the distribution (slope $\approx$ -0.43). The cases correspond to uniform sampling and extreme point sampling, labeled 'worst case'. The ranges of variation of the random input (bending stiffness D) are also indicated. Both figures demonstrate the concentration-of-measure phenomenon.

We begin by supposing that the model F is perfect. Fig. 4 depicts the dependence of the distribution in performance on the number of random inputs. The distributions are calculated directly by Monte Carlo sampling, with the sampling carried out in two ways: assuming a uniform distribution of the inputs over their intervals of variation, referred to as 'uniform distribution' in the figure; and assuming that the inputs are at the end points of their intervals of variation, referred to as 'worst case' in the figure. As expected, uniform sampling results in a lower probability of failure than worst-case sampling. All Monte Carlo calculations presented in the sequel are carried out using worst-case sampling. The standard deviation of the distribution is found to scale as $M^{-0.43}$ with the number of random variable inputs M, Fig. 4b. A marked concentration of the performance histogram as the number of inputs is increased is clearly evident in the figure, which vividly demonstrates the concentration-of-measure phenomenon on which the present approach is predicated.

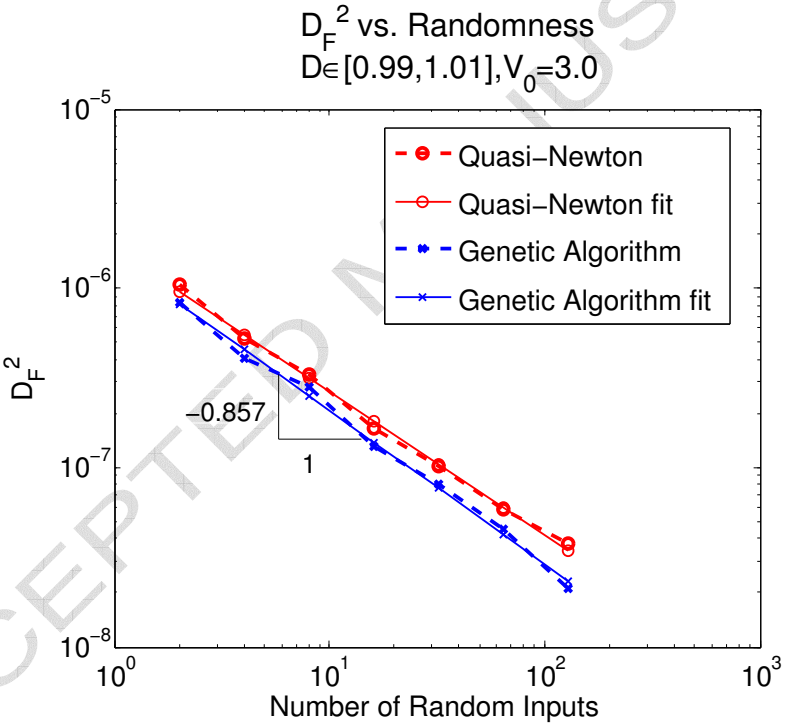


Figure 5: Ring implosion test. Dependence of the verification diameter (as computed by either a quasi-Newton iteration or a genetic algorithm iteration) on the number of random inputs of the system, showing clear power-law behavior.

The dependence of the verification diameter D_F of the response function on the number M of input parameters is shown in Fig. 5. For linear F one has the

scaling $D_F \sim M^{-\frac{1}{2}}$, and hence the power-law behavior ($D_F \sim M^{-0.4}$ in the quasi-Newton calculations, $D_F \sim M^{-0.43}$ in the genetic algorithm calculations) evident in Fig. 5 owes to the lack of regularity of the response function $F(X)$. This scaling behavior serves to illustrate one of the principal strengths of the concentration-of-measure approach, namely, that the uncertainty bounds become sharper as the number of random variables increases. Therefore, concentration-of-measure bounds are particularly attractive in the context of certification of large complex systems with many random inputs. It is also interesting to note from Fig. 5 that both the genetic algorithm and quasi-Newton results exhibit ostensibly identical behavior.

Figs. 6 and 7 show the minimum design margins required in order to guarantee that the concentration-of-measure bounds on the probability of failure are less than $\epsilon = 10^{-3}$, 10^{-2} and 10^{-1}. The results shown in Fig. 6 correspond to a perfect model whose mean performance is exactly known, termed *scenario 1* in Section 2.1, whereas the results shown in Fig. 7 correspond to a perfect model whose mean performance is estimated empirically, termed *scenario 3* in Section 2.3. The initial implosion velocity is $V_0 = 3$ in all cases. This initial velocity causes the ring to crumple extensively with the result that the behavior of the ring becomes exceedingly nonlinear. The figures show the dependence of the required margin on the number of random inputs for four input ranges: $D \in [0.95, 1.05]$, $[0.99, 1.01]$, $[0.995, 1.005]$ and $[0.999, 1.001]$. The figures also show a comparison of the margins resulting from aleatoric diameters D_F computed by a genetic algorithm and a Quasi-Newton iteration.

A number of features of the results presented in Figs. 6 and 7 immediately come to prominence. Interestingly, despite the extremely nonlinear behavior of the system the margins required to guarantee its safe performance take modest values which, presumably, should be attainable by practical designs. The modest range and good behavior of those design margins illustrates the feasibility and practicality of concentration-of-measure inequalities as a basis for the certification of complex systems. A second feature of interest is the steady decrease of the design margins with the number of random inputs, a decrease that is in keeping with the expected behavior of concentration-of-measure inequalities. Thus, the concentration-of-measure approach to certification pursued in the present work, which are based on the simple Hoeffding inequality, is particularly effective for systems with a large number of uncorrelated or weakly correlated random inputs. Extensions of concentration-of-measure inequalities to correlated inputs are summarized in Appendix A but will not be pursued here in the interest of simplicity. A final feature of interest, which is evident from a comparison of Figs. 6 and 7, concerns the small increase of the required design margins that results from estimating the mean performance of the system empirically. This small effect suggests

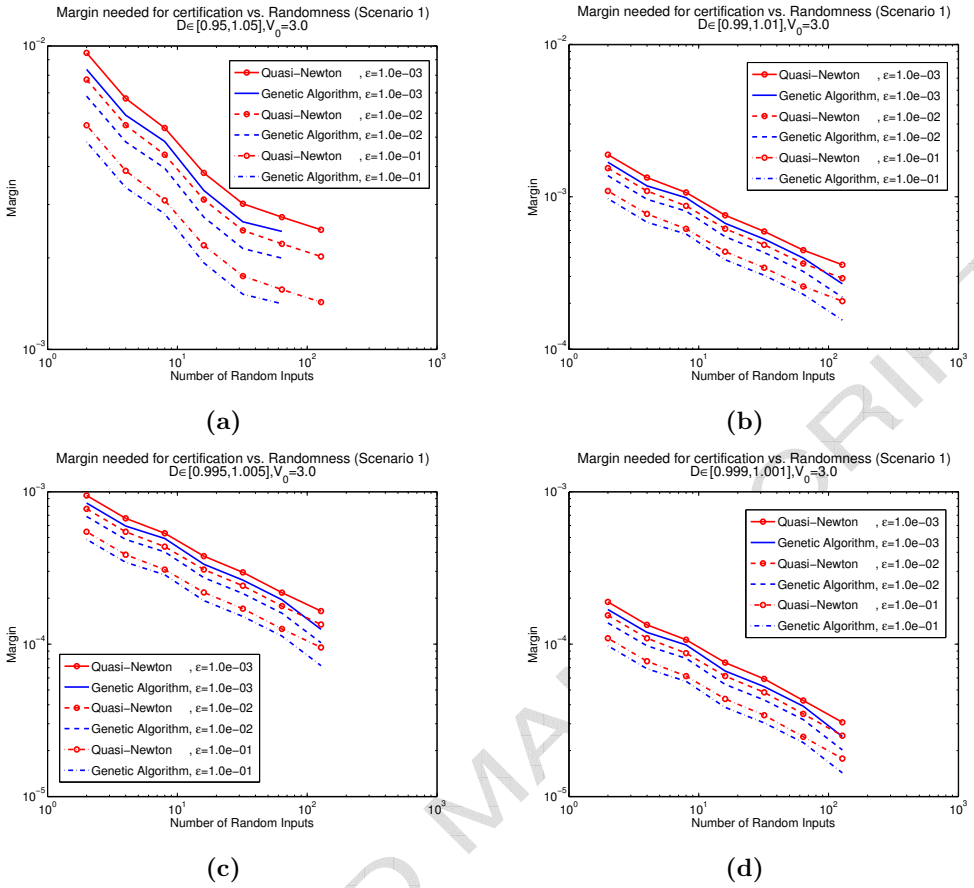


Figure 6: Minimum design margins required in order to guarantee probabilities of failure less than $\epsilon = 10^{-3}$, 10^{-2} and 10^{-1}. Perfect model and exact mean performance case. Initial implosion velocity $V_0 = 3$. The figures show the dependence of the required margin on the number of random inputs for four input ranges: $D = [0.95, 1.05]$, $[0.99, 1.01]$, $[0.995, 1.005]$ and $[0.999, 1.001]$. The figures also show comparison of the margins resulting from aleatoric diameters D_F computed by a genetic algorithm and a Quasi-Newton iteration.

that in practice mean performances can be safely estimated empirically and that the mean performance of the system need not be known exactly.

3.3.2 Inexact model

Next we investigate the effect of modeling uncertainties, including unknown unknowns. The precise manner in which the exact model is coarse-grained and unknown unknowns are introduced has been described in Section 3.1. The determination of the residual probability-of-failure tolerance requires the estimation

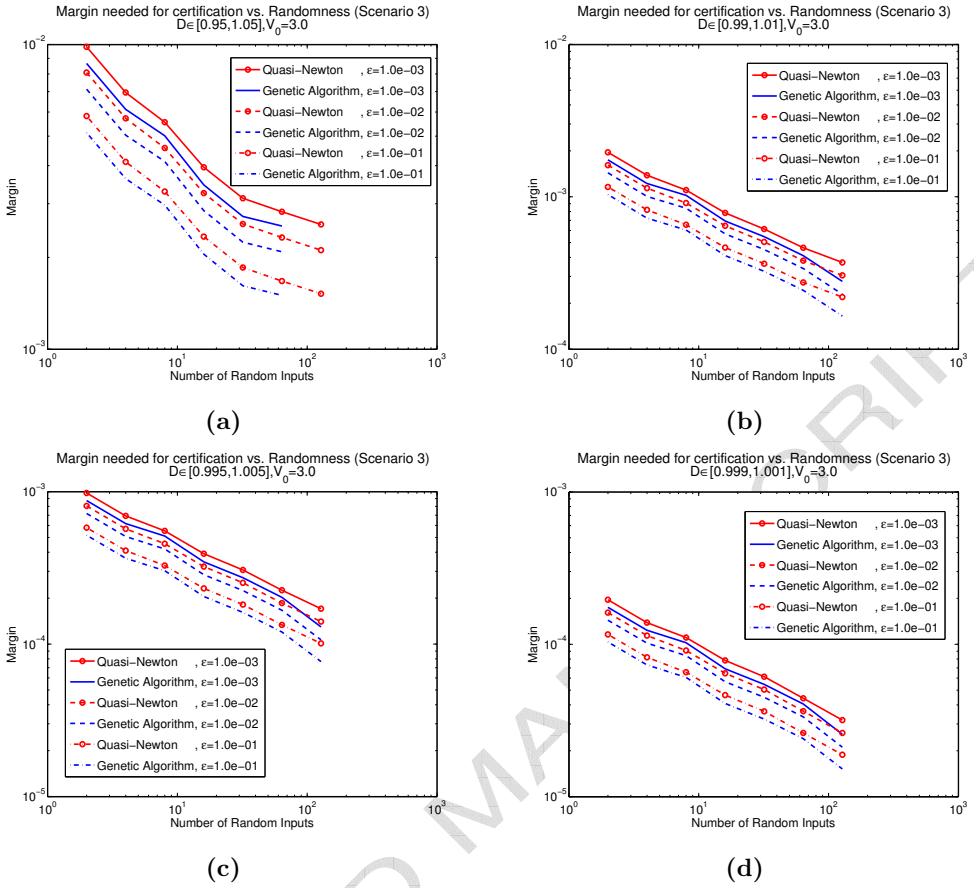


Figure 7: Minimum design margins required in order to guarantee probabilities of failure less than $\epsilon = 10^{-3}$, 10^{-2} and 10^{-1}. Perfect model and estimated mean performance case. Initial implosion velocity $V_0 = 3$. The figures show the dependence of the required margin on the number of random inputs for four input ranges: $D = [0.95, 1.05]$, $[0.99, 1.01]$, $[0.995, 1.005]$ and $[0.999, 1.001]$. The figures also show comparison of the margins resulting from aleatoric diameters D_F computed by a genetic algorithm and a Quasi-Newton iteration.

of the mean deviation $\langle G_i - F_i \rangle$ between predicted and measured performance measures and the computation of the validation diameters $D_{G_i - F_i}$. Again we note that this determination requires the simultaneous execution of nominally identical calculations and experiments, a process that gives precise form to the notion of validation.

As stated in Section 2.5, in the context of certification, models are useful precisely because, for sufficiently predictive models, the objective function required to compute the validation diameter, eq. (53), may be expected to exhibit much less variation than the response function itself, with the result that the evaluation of

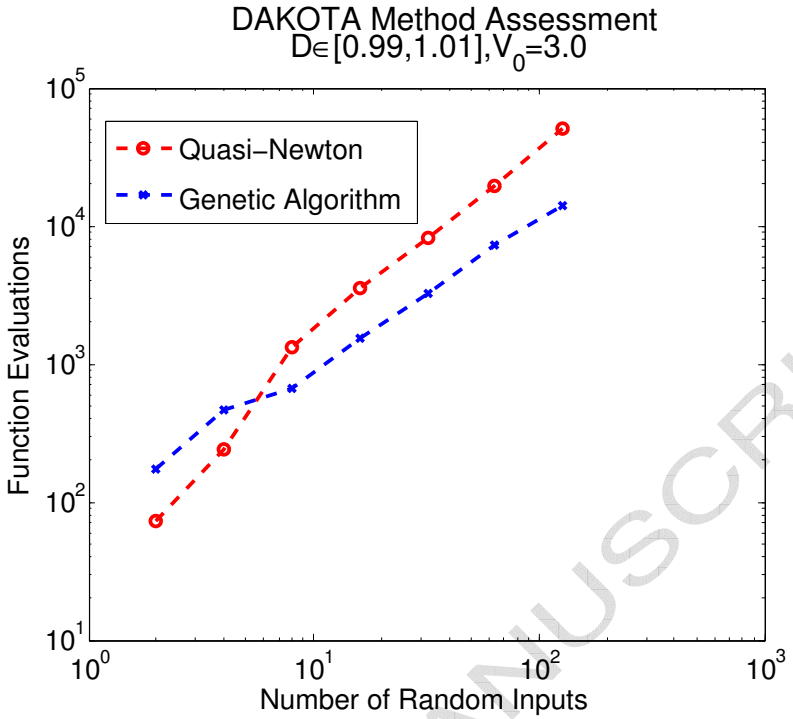


Figure 8: Ring implosion test. Computation of the validation diameter D_{G-F}. Comparison of the number of iterations to convergence using a quasi-Newton iteration and a genetic algorithm as function of the number of random input parameters. All calculations are carried out using Sandia National Laboratories' DAKOTA Version 4.0.

the validation diameter can be based on rapidly-convergent iterative procedures, which in turn can greatly reduce the number of experimental tests required for certification. This working assumption may be tested in the case of the imploding ring example. Fig. 8 shows plots of the number of iterations to convergence using a quasi-Newton iteration and a genetic algorithm as function of the number of random input parameters. As may be seen in the figure, for a small number of input parameters the quasi-Newton iteration converges more rapidly than the genetic algorithm and the number of iterations to convergence is manageably small. As the complexity of the system increases, the performance gap between the genetic algorithm and the quasi-Newton iteration narrow and, for a large number of input parameters the genetic algorithm requires fewer iterations to convergence.

In a certification context, the effect of the limited model fidelity is an effective reduction in the probability-of-failure tolerance, cf Section 2.5. Thus, the computed probability of failure must now be compared against a reduced tolerance that accounts for modeling errors. Fig. 9 shows that, for the particular choice of

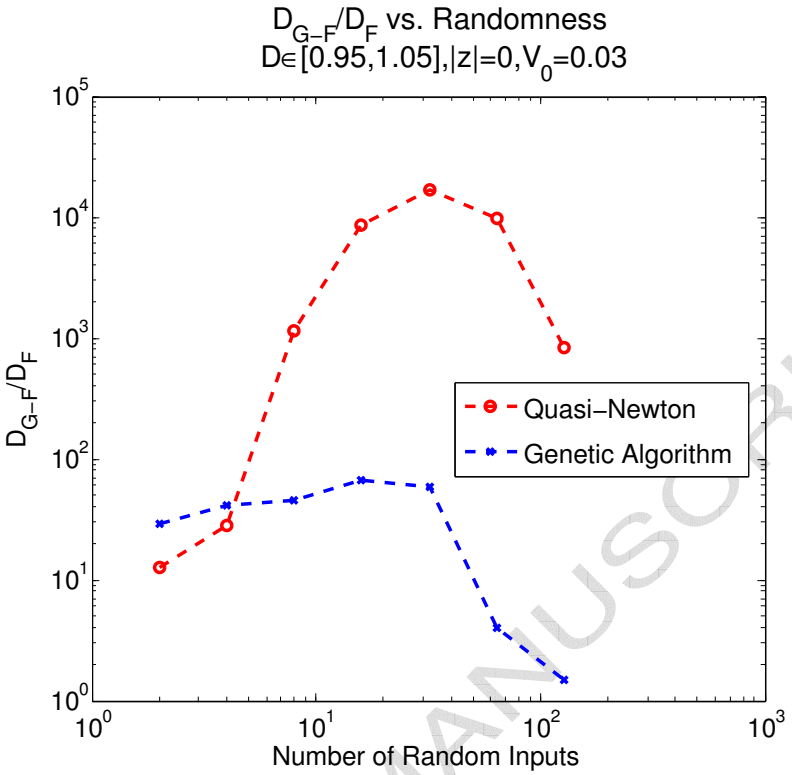
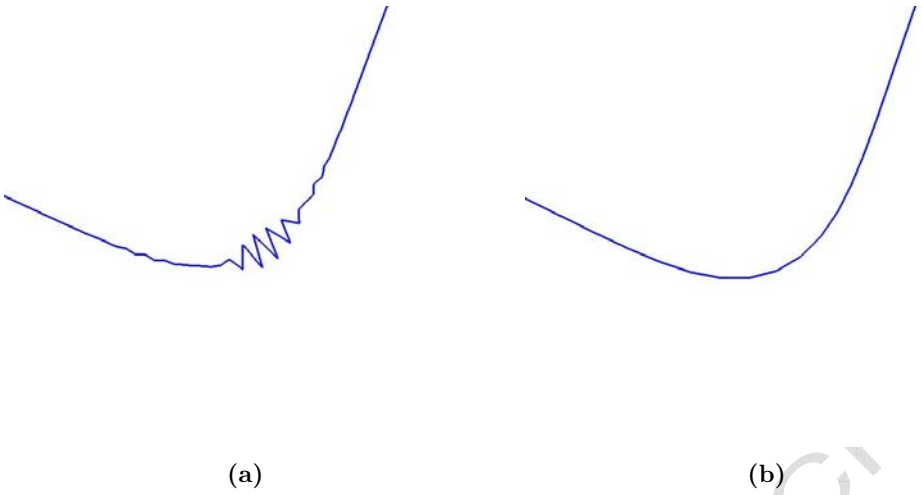


Figure 9: Ring implosion test. Ratio of validation diameter, D_{G-F}, to verification diameter, D_F.

upscaling model used in the calculations, the validation diameter D_{F-G} is greatly in excess of the verification diameter D_F. Thus, the epistemic or modeling uncertainty U_E completely overshadows aleatoric uncertainty U_A resulting from the intrinsic variability of the inputs and any stochastic nature of the system. The reason for the large epistemic uncertainty is clearly illustrated in Fig. 10, which shows that fine crumpling of the ring occurs in the late stages of compression, Fig. 10a. This localized crumpling greatly contributes to the bending energy of the ring, i. e., to the chosen measure of performance. However, the localized crumpling cannot be resolved by the coarse model and is completely suppressed, Fig. 10b, with the result that the bending energy of the ring is greatly underestimated.

In order to compensate for the large epistemic uncertainty, the certification of the system requires larger margins than otherwise required in the case of the perfect model, cf Fig. 11. This is particularly so for large numbers of random input parameters owing to the lack of concentration exhibited by the inexact model, a behavior that is in sharp contrast to the perfect model which, as noted earlier,



(a) **(b)**

Figure 10: Closeup of a snapshot of the ring trajectory showing: a) localized crumpling in the exact system represented by $G(X, Z)$; and b) the failure of the coarse-grained model $F(X)$ to resolved the localized crumpling.

exhibits strong concentration with increasing system size. The behavior of the system is not significantly altered by the introduction of unknown unknowns, cf Fig. 12. Remarkably, despite the large level of epistemic uncertainty and the introduction of unknown unknowns the minimum margins required for certification remain within relatively modest values attainable by practical designs. Thus, while the concentration advantage may be diminished, concentration-of-measure inequalities remain a viable avenue for certification in the context of inexact models and unknown unknowns.

4 Summary and Discussion

Certification is a process that seeks to establish whether the probability of failure of a system is below an acceptable tolerance. Often, certification is expressed in a language of quantification of *margins* and *uncertainties* (QMU), with the understanding that a system is certified if its performance uncertainty is less than its performance margin. Appealing as the QMU conceptual framework is, the precise quantitative definition of uncertainty measures, the precise means by which the uncertainty measures can be determined in practice, be it experimentally, by computer simulation or a combination of both, and the precise manner in which mitigating and controlling uncertainties *guarantees* the safe operation of a system, are often left unspecified.

We have developed a method of certification predicated upon the use of concen-

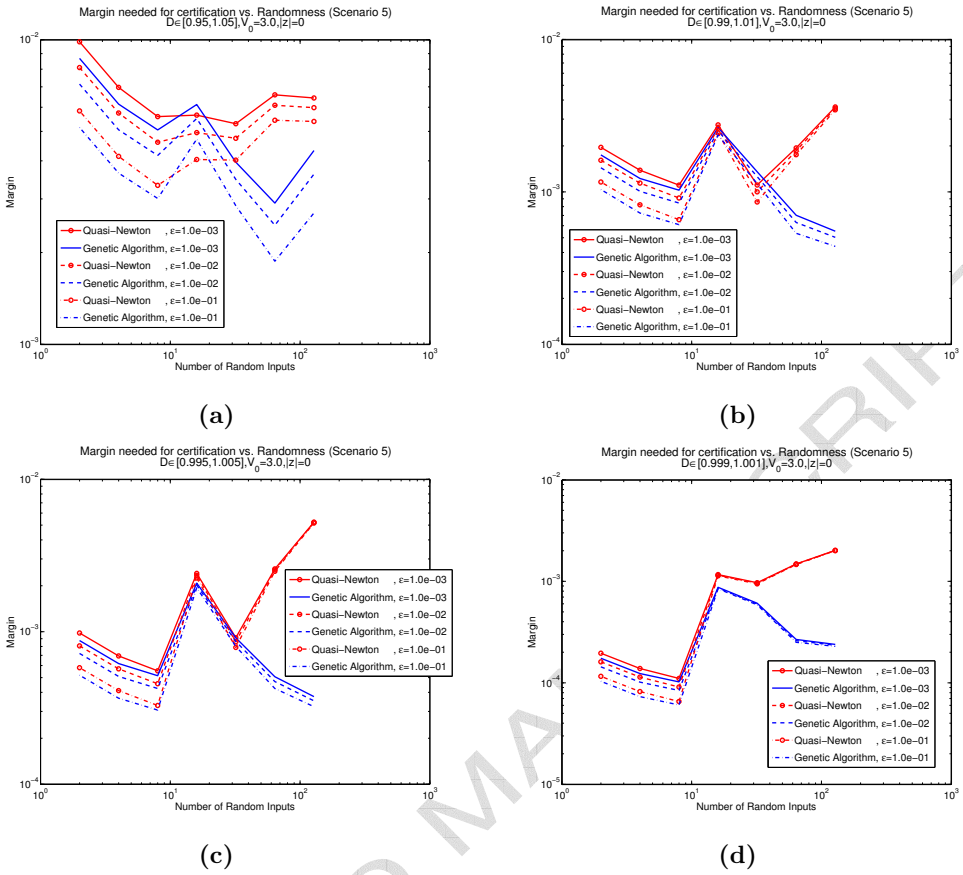


Figure 11: Minimum design margins required in order to guarantee probabilities of failure less than $\epsilon = 10^{-3}$, 10^{-2} and 10^{-1}. Inexact model, estimated mean performance and no unknown unknown case. Initial implosion velocity $V_0 = 3$. The figures show the dependence of the required margin on the number of random inputs for four input ranges: $D = [0.95, 1.05]$, $[0.99, 1.01]$, $[0.995, 1.005]$ and $[0.999, 1.001]$ for the coarse model F, with corresponding intervals for the fine model G obtained from the scaling relation (67b). The figures also show comparison of the margins resulting from aleatoric diameters D_F and epistemic diameters D_{G-F} computed by a genetic algorithm and a Quasi-Newton iteration.

tration-of-measure inequalities as a means of *bounding* performance uncertainties. These uncertainty bounds are mathematically rigorous and, therefore, can be taken as a basis for formulating conservative certification criteria. In addition, the approach is unambiguous and supplies precise quantitative definitions for a number of *terms of art* that are often loosely used in *ad hoc* certification methodologies, including the following:

i) *Margins*. When the mean performance of the system is known, the margin

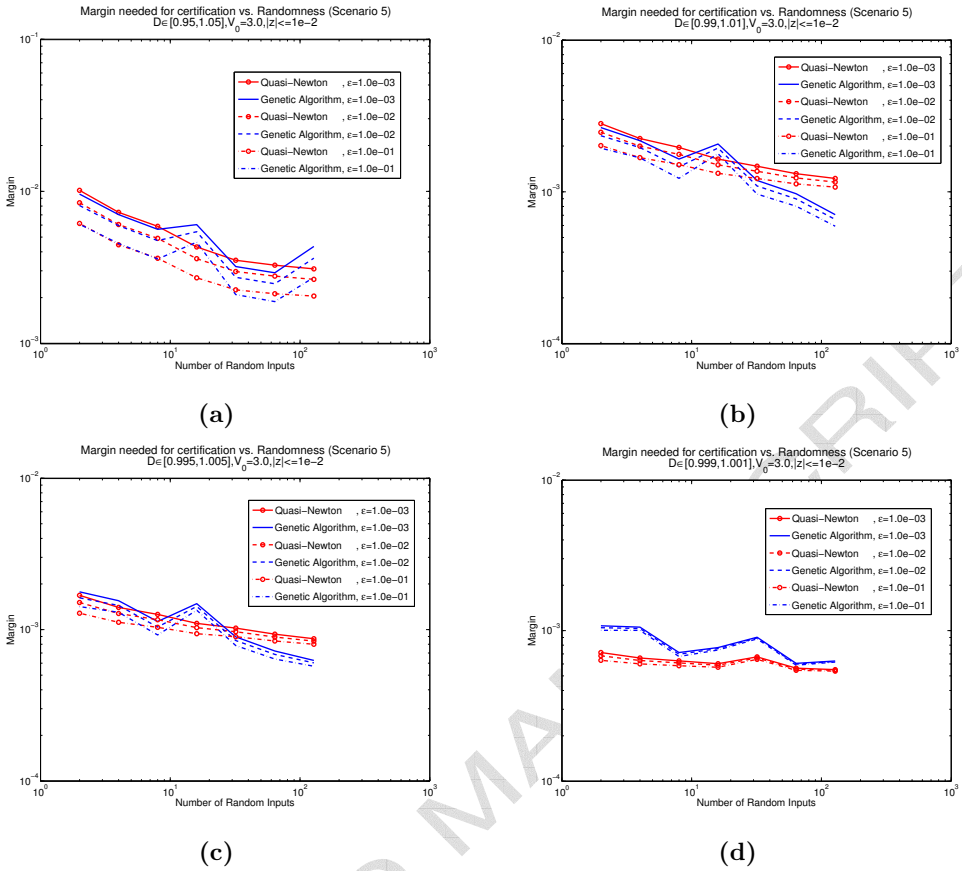


Figure 12: Minimum design margins required in order to guarantee probabilities of failure less than $\epsilon = 10^{-3}$, 10^{-2} and 10^{-1}. Inexact model and estimated mean performance. Unknown unknowns added to the system in the form of added bead masses resulting total ring mass variations in the range $[-0.1, 0.1]$. Initial implosion velocity $V_0 = 3$. The figures show the dependence of the required margin on the number of random inputs for four input ranges: $D = [0.95, 1.05]$, $[0.99, 1.01]$, $[0.995, 1.005]$ and $[0.999, 1.001]$ for the coarse model F, with corresponding intervals for the fine model G obtained from the scaling relation (67b). The figures also show comparison of the margins resulting from aleatoric diameters D_F and epistemic diameters D_{G-F} computed by a genetic algorithm and a Quasi-Newton iteration.

in a performance measure is simply the difference between the mean performance and its threshold for safe operation, eq. (11a). When the mean performance of the system is estimated by way of a model, the margin is the difference between the mean performance and an *increased* threshold, eq. (40a). This threshold increase accounts for uncertainties in the estimation of the mean performance and effectively results in a net loss of margin.

ii) *Uncertainty*. The total uncertainty of the system follows as the sum of an aleatoric uncertainty and an epistemic uncertainty, eq. (59b). The aleatoric uncertainty measures the spread in predicted performance arising from numerical errors, from the statistical variability of the input parameters, the intrinsic stochasticity of the model, or from other sources. The epistemic uncertainty measures the deviation between predicted and observed performance due to the limited fidelity of the model and existence of unknown unknowns.

iii) *Aleatoric uncertainty*. The aleatoric uncertainty in a performance measure is quantified by its verification diameter (8), i. e., the largest deviation in performance that is computed when each input parameter is allowed to vary in turn between pairs of values spanning its entire range. The aleatoric uncertainty is computed directly from the model without reference to experimental data. It bears emphasis that classical linearized *sensitivity analysis* is not sufficient to quantify aleatoric uncertainties in general. Instead, worst-case scenario *large deviations* in the system input parameters, resulting in likewise large deviations in system performance, must systematically be identified and taken into account. Finally, we note that aleatoric uncertainties can be determined on the sole basis of input parameter ranges without precise knowledge of their probability density functions.

iv) *Epistemic uncertainty*. The epistemic uncertainty in a performance measure is quantified by its validation diameter (53), i. e., the largest deviation in the difference between computed and measured performance that is recorded when each input parameter is allowed to vary in turn between pairs of values spanning its entire range. The determination of the validation diameter is an optimization problem in which the evaluation of the objective function requires the execution of nominally identical calculations and experiments. The algorithm employed in the solution of this optimization problem, be it a stochastic algorithm such as simulated annealing or a genetic algorithm or an iterative algorithm such as a quasi-Newton iteration, determines the precise sequence of calculations and experiments to be performed. It should be noted that, in executing *nominally identical* calculations and experiments, unknown-unknown input parameters are assigned random values within their respective ranges, which adds to the computed modeling uncertainties and systematically and automatically accounts for the effect of unknown unknowns.

v) *Confidence factor*. With the preceding definitions, the confidence factor is simply the quotient of margin to total uncertainty, eq. (10). However, it should be carefully noted that what is specifically asserted through this

definition is the mathematical fact that concentration-of-measure inequalities rigorously *guarantee* that a system whose confidence factor is above a well-defined threshold will operate safely within a pre-specified probability-of-failure tolerance. This stands in contrast to QMU methodologies based on *ad hoc* definitions of margins, uncertainties and confidence factors which, while eminently reasonable and intuitive appealing, may lack a similar mathematical guarantee and, therefore, fail provide a sound basis for certification.

vi) *Aggregation of uncertainties*. In cases in which the safe operation of a system requires multiple performance measures to be above their respective thresholds, certification can still be expressed in terms of an overall confidence factor for the system. In the concentration-of-measure approach to certification, the overall confidence factor follows as a well-defined function of the individual confidence factors of each of the performance measures, eq. (31). It bears emphasis that neither margins nor uncertainties can be aggregated in separation of each other. Instead, individual performance measure confidence factors, which naturally weigh individual margins against their corresponding uncertainties, are to be compounded into an overall confidence factor for the system. In particular, *ad hoc* formulae for aggregating uncertainties, such as root mean square (RMS) formulae, fail to provide a sound basis for certification in general.

It bears emphasis that, in a certification context, the purpose and utility of devising models of the highest possible fidelity is to minimize the number of—presumably costly—tests that are required for certification. *Predictive Science* may then be regarded as the art of formulating such models. Some of the benefits of the concentration-of-measure QMU framework towards the goal of achieving predictive science are worth noting carefully.

The computation of the verification and validation diameters of system requires a global optimization over parameter space. Global optimization algorithms such as simulated annealing and genetic algorithms exhibit a very high degree of concurrency, since at every step of the algorithms large populations of replicas of the model can be evaluated independently. This concurrent evaluation can be accomplished by running a large number of independent jobs on relatively small processor counts ($\sim$500-1000). Bottlenecks should be rare in this mode and exceedingly high efficiencies are expected. Petascale computing capacity presently under development will enable the running of vast numbers of such jobs simultaneously. Thus, concentration-of-measure uncertainty analysis lends itself ideally to—and provides a potentially important use of—petascale computing.

The calculation of aleatoric and epistemic uncertainties, as measured by the verification and validation diameters of the system, entails a systematic exploration of parameter space and thus identifies where the critical large-perturbation

sensitivities of the system and modeling errors reside. Thus, one important outcome of uncertainty quantification through the computation of the verification and validation diameters is the identification of *bad actors*, i. e., the components of the model, be they numerical, physics models or otherwise, responsible for the highest system uncertainties. Those model components can then be targeted for refinement through the addition of higher-fidelity physics laws, higher-accuracy numerical algorithms, or by other means. In this manner, QMU systematically guides model development and provides a rational basis for allocating modeling and experimental priorities and resources. Such systematic and prioritized model development is an integral and indispensable part of predictive science.

In closing, we proceed to enumerate some of the present limitations of the proposed approach and possible extensions and enhancements thereof. For simplicity, in this paper we have considered input parameters that are independent and supposed that all that is known about the variability in the input parameters is that they lie within certain intervals. However, the concentration-of-measure phenomenon is not limited to the case of independent random inputs and it is possible to devise concentration-of-measure inequalities that account for correlations between inputs (cf., e. g., theorems A3 and A4). It is also possible to devise concentration-of-measure inequalities that take as input probability density functions of the input variables, and take the resulting inequalities as a basis for formulating conservative certification criteria (cf., e. g., theorem A4).

An additional concern is whether the concentration-of-measure inequalities supply a sufficiently tight upper bound on the probability of failure. The test case of an imploding ring shows that the simplest concentration-of-measure inequalities, namely, those based on Hoeffding's inequality, can significantly overestimate the probability of failure. There are a number of alternative concentration-of-measure inequalities, such as Chernoff's inequality, that provide tighter upper bounds on the probability of failure. Another possibility is to exploit special characteristics of the system. For instance, a case that often arises in practice concerns systems that are composed of coupled *components* or *subsystems*. Each subsystem can then then be modeled separately and an integrated model of the entire system can subsequently be obtained by modeling the coupling between the subsystems. Hierarchies of models in which the subsystems can themselves be recursively decomposed into finer subsystems are also encountered in practice. In these cases, concentration inequalities can be applied recursively in order to bound uncertainties in the integrated system. The uncertainties in the subsystems can be computed through experiments tailored to each subsystem and through numerical simulations based on subsystem models, presumably an easier task than testing and modeling the integrated system itself. The resulting bounds on the subsystem uncertainties exhibit Gaussian tails. The uncertainty of subsystems higher in the

hierarchy can then be controlled by means of powerful concentration inequalities for Gaussian random variables. We recall that concentration inequalities having independent Gaussian random variables as input parameters return back Gaussian tail estimates. In this manner, uncertainties can be propagated recursively up the system hierarchy using concentration inequalities for Gaussian random variables.

It should be carefully noted that the present approach requires the determination of the epistemic uncertainties, as measured by the validation diameter, through the execution of a sequence of identical simulations and experiments. While this aspect of the method may be regarded as a strength—it supplies precise guidelines for model validation and leads to rigorous and conservative certification—it also limits the applicability of the method to systems for which integral tests can be conducted on demand. This raises the question of whether it is possible to extend the present approach—and, more generally, whether rigorous certification is possible at all—when only historical integral data is available and the possibility of acquiring new integral data does not exist. From a mathematical standpoint, the main difficulty is that in general there is no guarantee that historical integral data samples parameter space adequately, specially when certifying new designs, and that performance uncertainties are not underestimated as a result. The situation improves significantly if component testing is possible. Thus, if the system consists of a number of components and interfaces between the components, and if each component and interface can be tested, then it is possible to derive rigorous uncertainty bounds, leading to conservative certification, by methods similar to those outlined in the preceding paragraph.

Finally, we remark on the likely range of applicability of concentration-of-measure inequalities relative to other competing approaches. In cases in which the probability of failure is large, sampling methods such as Monte Carlo or Quasi Monte Carlo require a relatively small number of samples and are likely to enjoy a competitive advantage. Sampling methods are also advantageous when large data sets are available or are inexpensive to obtain. However, methods based on direct sampling become impractical if the probability of failure is small, i. e., if failure is a rare event, and if sampling is costly. By way of contrast, the effort required for the computation of the verification and validation diameters is independent of the size of the probability of failure and concentration-of-measure inequalities can conveniently be applied–and enjoy a competitive advantage–when probabilities of failure are small. Furthermore, concentration-of-measure inequalities are just about the only available certification tools for systems with a large number of inputs whose probability density function is not fully known.

Acknowledgements

The authors gratefully acknowledge the support of the Department of Energy through Caltech's ASCI ASAP Center for the Simulation of the Dynamic Response of Materials, and the support received from NSF through an ITR grant on Multiscale Modeling and Simulation and Caltech's Center for Integrative Multiscale Modeling and Simulation.

Appendix A. A brief review of concentration-of-measure inequalities

The concentration-of-measure phenomenon refers to the property that functions of a large number of variables, i. e., functions over high-dimensional spaces, with small local oscillations in each variable are almost constant. Moreover, the fluctuations can be controlled through powerful inequalities called concentration-of-measure inequalities. These inequalities have found a wide range of application in functional analysis, complexity theory, and probability and statistics but remain largely unknown outside the specialized mathematical community where they are currently being developed. In this appendix we collect basic facts and background on the concentration-of-measure phenomenon for the convenience of the reader. We refer the interested reader to [6] for a monograph and to [7] for a survey.

Recognition of the concentration-of-measure phenomenon as such may be traced back to an observation by Lévy [13] that functions on high dimensional spheres with small local oscillations, i. e., whose modulus of continuity can be controlled, are strongly concentrated around their mean value with respect to the uniform Lebesgue measure on the hypersphere. The study of this phenomenon was pioneered in the early seventies by V. Milman in his work on the asymptotic geometry of Banach spaces [14], [15], [16], [17]. For an isoperimetric interpretation of the concentration-of-measure phenomenon and powerful applications to geometry we refer to [18], [19] and [20]. For early probabilistic results in the context of sums of independent random variables we refer to [9] (Hoeffding's inequality), [21] (Chernoff bound) and to ([22] and [23]) for quantitative inequalities pertaining to the Glivenko-Cantelli [24], [25] convergence of empirical distributions. Far-reaching extensions, that in particular provide dimension-free concentration-of-measure inequalities in product spaces, have more recently been advanced by M. Talagrand, cf. [26], [27], [28], [29], [30], [31], [32]. For a selection of relevant articles in probability and statistics we refer to [33], [34], [35], [36], [37], [38], [39] and [40].

A brief compendium of representative concentration-of-measure inequalities is collected in the following. Whereas only the simplest McDiarmid concentration-of-measure inequalities is used in this paper, the more advanced inequalities de-

scribed below supply avenues for extension of the present QMU methodology to systems including correlated inputs, inputs with known probability distributions, unbounded inputs and other cases of interest.

McDiarmid's inequality. *McDiarmid's inequality* is perhaps the simplest example of concentration-of-measure inequality. Let $X_1, \ldots, X_M$ be M random variables with values in spaces $E_1, \ldots, E_M$. Let F be a-one dimensional function of $X_1, \ldots, X_M$. Write

$$D_F^2 := \sum_{i=1}^{M} \sup_{(x_1,\ldots,x_{i-1},x_{i+1},\ldots x_M)\in E_1\times\cdots E_{i-1}\times E_{i+1}\times\cdots\times E_M} \sup_{(A_i,B_i)\in E_i^2} |F(x_1,\ldots,x_{i-1},A_i,x_{i+1},\ldots x_M) - F(x_1,\ldots,x_{i-1},B_i,x_{i+1},\ldots x_M)|^2. \tag{A1}$$

We refer to D_F as the diameter of F. In Eq. (A1) the suprema are taken with respect to variables in the spaces $E_1,\ldots, E_M$. Let $\mathbb{P}$ and $\mathbb{E}$ be the measure of probability and expectation associated with the random variables $X_1, \ldots, X_M$. Write $X := (X_1, \ldots, X_M)$. The following theorem [10], also known as the *bounded-differences inequality*, bounds the fluctuations of $F(X)$ away from its mean without *a priori* knowledge of the probability distribution of the random variables $X_1, \ldots, X_M$.

Theorem A1 (McDiarmid, 1989) *Let the random variables $X_1, \ldots, X_M$ are independent. Then*

$$\mathbb{P}[F(X) - \mathbb{E}[F] \geq r] \leq \exp(-2\frac{r^2}{D_F^2}) \tag{A2}$$

Observe that if the spaces $E_i = (a_i, b_i)$ and

$$F(X) = \frac{1}{M}\sum_{i=1}^{M} X_i\,, \tag{A3}$$

Hoeffding's inequality [9]

$$\mathbb{P}[\frac{1}{M}\sum_{i=1}^{M} X_i - \frac{1}{M}\sum_{i=1}^{M}\mathbb{E}[X_i] \geq r] \leq \exp(-2M\frac{r^2}{(\sum_{i=1}^{M}(b_i-a_i)^2/M)}) \tag{A4}$$

is recovered as a special case.

Convex-distance inequality. The bounded-differences inequality is a special case of the more powerful *convex-distance inequality* [30]. In particular, the

convex-distance inequality applies in cases where the bounded-differences inequality fails. Assume that $X_1, \ldots, X_M$ are independent variables, each taking values in a measurable set E. Write $X = (X_1, \ldots, X_M)$ and $\mathbb{P}[A] := \mathbb{P}[X \in A]$. Let $A \subset E^M$ be an arbitrary measurable subset of E. For $\alpha \in [0,1]^M$, we define the weighted Hamming distance from the point $x \in E^M$ to A as

$$d_\alpha(x, A) := \inf_{z \in A} \sum_{i:\, z_i \neq x_i} |\alpha_i| \tag{A5}$$

Writing $\|\alpha\|^2 := \sum_{i=1}^M \alpha_i^2$ we define the convex distance of X from the set A as

$$d_T(x, A) := \sup_{\alpha \in [0,1]^M, \|\alpha\|=1} d_\alpha(x, A) \tag{A6}$$

Then we have the following.

Theorem A2 *For any subset $A \subset E^M$ with $\mathbb{P}[X \in A] \geq \frac{1}{2}$ and $t > 0$,*

$$\min\big\{\mathbb{P}[A], \mathbb{P}[d_T(X, A) \geq t]\big\} \leq \exp(-\frac{t^2}{4}) \tag{A7}$$

The convex-distance inequality originates in a remarkable series of papers by Talagrand [27], [30], [41]. The preceding statement of the inequality is taken from [42].

Concentration inequalities with correlated random variables. The concentration-of-measure phenomenon is not limited to the case of independent random inputs and also arises when the inputs are correlated. Suppose, for definiteness, that $E_i = [0,1]$ for all i. Let $f\big(X_1, \ldots, X_M\big)$ be the probability density of the inputs and denote by $f\big(X_j, \ldots, X_M | (X_1, \ldots, X_{i-1}, X_i) = (x_1, \ldots, x_{i-1}, x_i)\big)$ the the law of $X_j, \ldots, X_M$ conditioned on $(X_1, \ldots, X_{i-1}, X_i) = (x_1, \ldots, x_{i-1}, x_i)$. A matrix Γ measuring the correlations between pairs of random variables X_i may then be defined as follows: $\Gamma_{ij} := 0$ if $i > j$; $\Gamma_{ii} := 1$; and, for $i < j$,

$$\begin{aligned} \Gamma_{ji} = \sup_{x_i, z_i \in [0,1]^2} \sup_{(x_1, \ldots, x_{i-1}) \in [0,1]^{i-1}} \\ \|f(X_j, \ldots, X_M | (X_1, \ldots, X_{i-1}, X_i) = (x_1, \ldots, x_{i-1}, x_i)) - \\ f(X_j, \ldots, X_M | (X_1, \ldots, X_{i-1}, X_i) = (x_1, \ldots, x_{i-1}, z_i))\|_{TV}\,, \end{aligned} \tag{A8}$$

where $\|.\|_{TV}$ denotes the total variation norm over probability measures. For instance, if $(X_i)_{1 \leq i \leq M}$ is a Markov Chain on $[0,1]$ with uniformly contracting transition kernels, i. e.,

$$\alpha := \sup_{i, x_{i-1}, z_{i-1}} \|f(X_i | X_{i-1} = x_i) - f(X_i | X_{i-1} = z_i)\|_{TV} < 1\,, \tag{A9}$$

then

$$\|\Gamma\| \leq \frac{1}{1-\alpha^{\frac{1}{2}}}\,. \tag{A10}$$

Write $\|\Gamma\|$ for the operator norm of the matrix Γ. Then we have the following.

Theorem A3 *Let F be 1-Lipschitz over $[0,1]^M$. Then*

$$\mathbb{P}[|F-\mathbb{E}[F]| \geq r] \leq 2\exp\big[-\frac{r^2}{2\|\Gamma\|^2}\big] \tag{A11}$$

We refer to [43] for a proof of the following theorem, and to to [44], [45], [37], [46] for related results.

Concentration inequalities with functions with unbounded oscillations. Concentration-of-measure inequalities are not limited to functions with bounded differences or inputs taking their values on compact spaces. General concentration-of-measure inequalities can be obtained by controlling the Lipschitz regularity of the output function and the tail of the random variables X_i at infinity. This control can be achieved by means of analytical inequalities called logarithmic Sobolev inequalities [47]. Let f be a non-negative measurable function f over a measure space $(E,\mathcal{B},\mu)$ such that $\int f\ln(1+f)d\mu<\infty$. We define the entropy of f as

$$\mathrm{Ent}_\mu(f) := \int f\ln f d\mu - \int f d\mu \ln(\int f d\mu)\,. \tag{A12}$$

We refer to Section 5 of [6] for the following theorem and to references therein.

Theorem A4 *Let μ be a probability measure on the Borel sets of a metric space (E,d) such that for some $C>0$ and all locally Lipschitz function f on E*

$$\mathrm{Ent}_\mu(f^2) \leq 2C\int |\nabla f|^2 d\mu \tag{A13}$$

with

$$|\nabla f|(x) := \limsup_{y\to x}\frac{|f(x)-f(y)|}{|x-y|}\,. \tag{A14}$$

Then for every 1-Lipschitz integrable function $F: E\to R$ and for every $r\geq 0$

$$\mu\big(\{Y\geq \int F d\mu + r\}\big) \leq \exp\big(-\frac{r^2}{2C}\big)\,. \tag{A15}$$

Eq. (A13) is called logarithmic Sobolev inequality with constant C. The application of theorem A4 to product of metric spaces (E_i,d_i) follows from the observation that if the measures μ_i satisfy the logarithmic Sobolev inequality

$$\mathrm{Ent}_{\mu_i}(f^2) \leq 2C_i\int |\nabla_i f|^2 d\mu_i \tag{A16}$$

for every locally Lipschitz function f on E_i, where $|\nabla_i f|$ is the generalized modulus of gradient on E_i defined as in Eq. (A14), then the product measure $\mu = \mu_1 \times \cdots \times \mu_M$ satisfies the logarithmic inequality (see corollary 5.7 of [6])

$$\mathrm{Ent}_\mu(f^2) \leq 2 \max_{1\leq i\leq M} C_i \int |\nabla f|^2 d\mu \tag{A17}$$

with $|\nabla f|^2 = \sum_{i=1}^M |\nabla_i f|^2$, for every locally Lipschitz function f on $E = E_1 \times \cdots \times E_M$. The particular application to Gaussian distributions follows simply from the observation that a normal centered Gaussian distribution on $\mathbb{R}^n$ satisfies the logarithmic Sobolev inequality with constant 2.

Concentration inequalities for empirical processes. Concentration inequalities can also be used to obtain powerful and very quantitative estimates for empirical processes defined by sampling, [27], [30] and [41] and [35]. Let $Y^1, \ldots, Y^N$ be independent random variables, not necessarily identically distributed, in some measurable space $(E, \mathcal{B})$. Let $\mathcal{F}$ be some countable family of real-valued measurable functions on $(E, \mathcal{B})$ such that $\|f\|_\infty \leq b < \infty$ for every $f \in \mathcal{F}$. Let

$$Z := \sup_{f\in\mathcal{F}} \Big| \sum_{i=1}^N f(Y_i) \Big| \tag{A18}$$

or

$$Z := \sup_{f\in\mathcal{F}} \Big| \sum_{i=1}^N (f(Y_i) - \mathbb{E}[f(Y_i)]] \Big| \tag{A19}$$

Let, in addition,

$$\sigma^2 := \sup_{f\in\mathcal{F}} \sum_{i=1}^N \mathrm{Var}[f(Y_i)] \tag{A20}$$

Then we have the following [35].

Theorem A5 *For any positive real number ϵ and x,*

$$\mathbb{P}\Big[Z \geq (1+\epsilon)\mathbb{E}[Z] + \sigma\sqrt{2\kappa x} + \kappa(\epsilon) b x \Big] \leq \exp(-x) \tag{A21}$$

with $\kappa = 4$, $\kappa(\epsilon) = 2.5 + 32\epsilon^{-1}$. Moreover,

$$\mathbb{P}\Big[Z \leq (1-\epsilon)\mathbb{E}[Z] - \sigma\sqrt{2\kappa' x} - \kappa'(\epsilon) x \Big] \leq \exp(-x) \tag{A22}$$

with $\kappa' = 5.4$ and $\kappa'(\epsilon) = 2.5 + 43.2\epsilon^{-1}$.

In the particular case in which the random variables Y_i are identically distributed, theorem A5 furnishes powerful estimates for the empirical process

$$\mu_N := \frac{1}{N}\sum_{i=1}^{N}\delta_{Y_i} \tag{A23}$$

by observing that

$$Z = \sup_{f\in\mathcal{F}} N|\mu_N(f) - \mu(f)| \tag{A24}$$

References

[1] D. H. Sharp and M. M. Wood-Schultz. QMU and nuclear weapons certification: What's under the hood. Los Alamos Science, (28):47–53, 2003.

[2] D. Eardley (Study Leader). Quantification of margins and uncertainties (QMU). Technical Report JSR-04-330, JASON, The MITRE Corporation, 7515 Colshire Drive McLean, Virginia 22102, March 2005.

[3] M. Pilch, T. G. Trucano, and J. C. Helton. Ideas underlying quantification of margins and uncertainties (QMU): A white paper. Unlimited Release SAND2006-5001, Sandia National Laboratory, Albuquerque, New Mexico 87185 and Livermore, California 94550, September 2006.

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

[5] W. L. Oberkampf, T. G. Trucano, and C. Hirsch. Verification, validation and predictive capability in computational engineering and physics. Applied Mechanics Reviews, 57(5):345–384, 2004.

[6] Michel Ledoux. The concentration of measure phenomenon, volume 89 of Mathematical Surveys and Monographs. American Mathematical Society, Providence, RI, 2001.

[7] S. Boucheron, O. Bousquet, and G. Lugosi. Concentration inequalities. In O. Bousquet, U.v. Luxburg, and G. Rtsch, editors, Advanced Lectures in Machine Learning, pages 208–240. Springer, 2004.

[8] Lawrence C. Evans. An introduction to stochastic differential equations. 2004. available at http://math.berkeley.edu/∼evans/SDE.course.pdf.

[9] Wassily Hoeffding. Probability inequalities for sums of bounded random variables. J. Amer. Statist. Assoc., 58:13–30, 1963.

[10] Colin McDiarmid. On the method of bounded differences. In Surveys in combinatorics, 1989 (Norwich, 1989), volume 141 of London Math. Soc. Lecture Note Ser., pages 148–188. Cambridge Univ. Press, Cambridge, 1989.

[11] Colin McDiarmid. Centering sequences with bounded differences. Combin. Probab. Comput., 6(1):79–86, 1997.

[12] T. J. R. Hughes. Analysis of transient algorithms with particular reference to stability behavior. In T. Belytschko and T. J. R. Hughes, editors, Computational Methods for Transient Analysis, pages 67–155. North-Holland, 1983.

[13] Paul Lévy. Problèmes concrets d'analyse fonctionnelle. Avec un complément sur les fonctionnelles analytiques par F. Pellegrino. Gauthier-Villars, Paris, 1951. 2d ed.

[14] V. D. Milman. Geometric theory of Banach spaces. II. Geometry of the unit ball. Uspehi Mat. Nauk, 26(6(162)):73–149, 1971.

[15] V. D. Milman. A certain property of functions defined on infinite-dimensional manifolds. Dokl. Akad. Nauk SSSR, 200:781–784, 1971.

[16] V. D. Milman. Asymptotic properties of functions of several variables that are defined on homogeneous spaces. Dokl. Akad. Nauk SSSR, 199:1247–1250, 1971.

[17] V. D. Milman. A new proof of A. Dvoretzky's theorem on cross-sections of convex bodies. Funkcional. Anal. i Priložen., 5(4):28–37, 1971.

[18] M. Gromov. Isoperimetry of waists and concentration of maps. Geom. Funct. Anal., 13(1):178–215, 2003.

[19] M. Gromov. CAT(κ)-spaces: construction and concentration. Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI), 280(Geom. i Topol. 7):100–140, 299–300, 2001.

[20] M. Gromov and V. D. Milman. Brunn theorem and a concentration of volume phenomena for symmetric convex bodies. In Israel seminar on geometrical aspects of functional analysis (1983/84), pages V, 12. Tel Aviv Univ., Tel Aviv, 1984.

[21] Herman Chernoff. A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. Ann. Math. Statistics, 23:493–507, 1952.

[22] N. V. Smirnov. Approximate laws of distribution of random variables from empirical data. Uspehi Matem. Nauk, 10:179–206, 1944.

[23] A. Dvoretzky, J. Kiefer, and J. Wolfowitz. Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. Ann. Math. Statist., 27:642–669, 1956.

[24] F Cantelli. Sulla probabilita come limita della frequenza. Rend. Accad. Lincei, 26(1), 1933.

[25] V Glivenko. Sulla determinazione empirica delle leggi di probabilita. Giornale dell'Istituta Italiano degli Attuari, 4, 1933.

[26] Michel Talagrand. Concentration and influences. Israel J. Math., 111:275–284, 1999.

[27] Michel Talagrand. New concentration inequalities in product spaces. Invent. Math., 126(3):505–563, 1996.

[28] Michel Talagrand. Nouvelles inégalités de concentration à q points. C. R. Acad. Sci. Paris Sér. I Math., 321(11):1505–1507, 1995.

[29] Michel Talagrand. Nouvelles inégalités de concentration "convexifiées". C. R. Acad. Sci. Paris Sér. I Math., 321(10):1367–1370, 1995.

[30] Michel Talagrand. Concentration of measure and isoperimetric inequalities in product spaces. Inst. Hautes Études Sci. Publ. Math., (81):73–205, 1995.

[31] Michel Talagrand. A new isoperimetric inequality and the concentration of measure phenomenon. In Geometric aspects of functional analysis (1989–90), volume 1469 of Lecture Notes in Math., pages 94–124. Springer, Berlin, 1991.

[32] WanSoo T. Rhee and Michel Talagrand. A concentration inequality for the K-median problem. Math. Oper. Res., 14(2):189–202, 1989.

[33] Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. Concentration inequalities using the entropy method. Ann. Probab., 31(3):1583–1614, 2003.

[34] Pascal Massart. Some applications of concentration inequalities to statistics. Ann. Fac. Sci. Toulouse Math. (6), 9(2):245–303, 2000. Probability theory.

[35] Pascal Massart. About the constants in Talagrand's concentration inequalities for empirical processes. Ann. Probab., 28(2):863–884, 2000.

[36] Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. A sharp concentration inequality with applications. Random Structures Algorithms, 16(3):277–292, 2000.

[37] Michel Ledoux. Concentration of measure and logarithmic Sobolev inequalities. In Séminaire de Probabilités, XXXIII, volume 1709 of Lecture Notes in Math., pages 120–216. Springer, Berlin, 1999.

[38] S. Bobkov and M. Ledoux. Poincaré's inequalities and Talagrand's concentration phenomenon for the exponential distribution. Probab. Theory Related Fields, 107(3):383–400, 1997.

[39] M. Ledoux. A heat semigroup approach to concentration on the sphere and on a compact Riemannian manifold. Geom. Funct. Anal., 2(2):221–224, 1992.

[40] M. Ledoux. A remark on hypercontractivity and the concentration of measure phenomenon in a compact Riemannian manifold. Israel J. Math., 69(3):361–370, 1990.

[41] Michel Talagrand. A new look at independence. Ann. Probab., 24(1):1–34, 1996.

[42] Gábor Lugosi. Concentration-of-measure inequalities. Lecture notes., 2006. http://www.econ.upf.edu/∼lugosi/anu.pdf.

[43] Paul-Marie Samson. Concentration of measure inequalities for Markov chains and Φ-mixing processes. Ann. Probab., 28(1):416–461, 2000.

[44] K. Marton. A measure concentration inequality for contracting Markov chains. Geom. Funct. Anal., 6(3):556–571, 1996.

[45] Katalin Marton. Measure concentration for a class of random processes. Probab. Theory Related Fields, 110(3):427–439, 1998.

[46] C. Houdré and P. Tetali. Concentration of measure for products of Markov kernels and graph products via functional inequalities. Combin. Probab. Comput., 10(1):1–28, 2001.

[47] Leonard Gross. Logarithmic Sobolev inequalities. Amer. J. Math., 97(4):1061–1083, 1975.