Applied Mathematical Modelling 54 (2018) 710–721

Contents lists available at ScienceDirect

Applied Mathematical Modelling

journal homepage: www.elsevier.com/locate/apm

Do we have enough data? Robust reliability via uncertainty quantification

Roberto Rocchetta [a], Matteo Broggi [b], Edoardo Patelli [a,*]

[a] Institute of Risk and Uncertainty, University of Liverpool, L69 3GQ Liverpool, United Kingdom
[b] Institute for Risk and Reliability, Leibniz Hannover University, Hannover 30167, Germany

ARTICLE INFO

Article history:
Received 5 December 2016
Revised 7 October 2017
Accepted 18 October 2017
Available online 26 October 2017

Keywords:
Uncertainty quantification
Information quality
Probability boxes
Dempster–Shafer
Computational tool
Reliability

ABSTRACT

A generalised probabilistic framework is proposed for reliability assessment and uncertainty quantification under a lack of data. The developed computational tool allows the effect of epistemic uncertainty to be quantified and has been applied to assess the reliability of an electronic circuit and a power transmission network. The strength and weakness of the proposed approach are illustrated by comparison to traditional probabilistic approaches. In the presence of both aleatory and epistemic uncertainty, classic probabilistic approaches may lead to misleading conclusions and a false sense of confidence which may not fully represent the quality of the available information. In contrast, generalised probabilistic approaches are versatile and powerful when linked to a computational tool that permits their applicability to realistic engineering problems.

1. Introduction

Nowadays it is generally well accepted that estimating the effect of uncertainty is a necessity, e.g. due to variation in parameters, operational conditions and in the modelling and simulations [1,2]. In practical applications, situations are common where the analyst has to deal with poor quality data, few available specimens or inconsistent information. A typical example is a situation where very expensive samples have to be collected, such as field proprieties of a deep reservoir [3] or performance of satellites [4]. In these cases, the amount of data will be scarce due to economic and time constraints and in several cases, expert elicitation (i.e. the best estimate of an expert) may be the only viable way of carrying on with the analysis [5].

As a consequence, strong assumptions may be needed to apply classical probabilistic methods given poor information quality, which can lead to erroneous reliability estimations and a false sense of confidence [6]. Generalised approaches, which fit in the framework of imprecise probability [6], are powerful methodologies for dealing with imprecise information and lack of data. These methodologies can be coupled to traditional probabilistic approaches in order to give a different prospective on the results, whilst avoiding the inclusion of unjustified assumptions and enhancing the overall robustness of the analysis. Generalised methods are rarely used in practice and this is probably due to lack of proper guidance, simulation tools, as well as some misconception in the interpretation of the results. Further comparison of different methodologies, both in theoretical aspects and in their applicability to real case studies, are required.

* Corresponding author.
E-mail addresses: Roberto.Rocchetta@liverpool.ac.uk (R. Rocchetta), broggi@bauinf.uni-hannover.de (M. Broggi), Edoardo.Patelli@liverpool.ac.uk (E. Patelli).

https://doi.org/10.1016/j.apm.2017.10.020
0307-904X/

An original throughout analysis of the applicability of different methodologies to deal with different level of imprecision is presented. In addition, this paper presents a novel computational framework for generalised probabilistic analysis that can be adopted to deal with low quality data, few available samples and inconsistent information. Efficient and generally applicable computational strategies have been developed and implemented into OpenCossan [7]. The proposed framework is applied to assess the reliability of an electric series RLC circuit (a problem proposed by the NAFEMS Stochastics Working Group [8]) and of a power transmission network, both affected by a lack of data.

Generally speaking, different system performance indicators may be affected very differently by the same (lack of) data. The extent of a lack of information is not a-priori quantifiable and depends on the context of the analysis. The proposed approach is used to assess the information quality by comparison to classical probabilistic results and with respect to system reliability estimates. One of the main contributions of this work is a detailed comparison between classical and generalised probabilistic approaches from a straightforward applicative point of view and under different levels of imprecision. This serves as guidance for engineering practitioners to solve problems affected by a lack of data.

The rest of the paper is structured as follows: Section 2 presents the mathematical framework. In Section 3, a synthetic overview of the numerical framework and the proposed approach is proposed. The NAFEMS reliability problem is described and solved in Section 4. A lack of data problem for power network reliability estimation is solved in Section 5. A discussion on the limitations of the different approaches is presented in Sections 6 and7 closes the paper.

2. Mathematical framework

Uncertainty is generally classified into two categories, aleatory and epistemic uncertainty. Aleatory uncertainty (Type I or irreducible uncertainty), represents stochastic behaviours and randomness of events and variables. Hence, due to its intrinsic random nature it is normally regarded as irreducible. Some examples of aleatory uncertainty are future weather conditions, stock market prices or chaotic phenomenon. Epistemic uncertainty (Type II or reducible uncertainty), is commonly associated with lack of knowledge about phenomena, imprecision in measurements and poorly designed models. It is considered to be reducible since further data can decrease the level of uncertainty, but this might not always be practical or feasible. In recent decades, efforts were focused on the explicit treatment of imprecise knowledge, non-consistent information and both epistemic and aleatory uncertainty. The methodologies are discussed in literature by different mathematical concepts: Evidence theory [9], interval probabilities [10], Fuzzy-based approaches [11], info-gap approaches [12] and Bayesian frameworks [13] are some of the most intensively applied concepts.

In this paper, Dempster-Shafer structures and probability boxes are used to model quantities affected by epistemic uncertainty, by aleatory uncertainty, or by a combination of the two. In addition, the Kolmogorov–Smirnov test [14] and Kernel Density Estimator [15] have been used to characterise the parameter uncertainty in case of small sample sizes.

2.1. Dempster–Shafer structures and probability boxes

The Dempster–Shafer (DS) theory is a well-suited framework to represent both aleatory and epistemic uncertainty. The difference between the axioms of classical probability theory and the DS theory is that the latter slacken the strict assumption of a single probability measure for an event. It can be seen as a generalisation of Bayesian probability [16]. Mathematically, a Dempster–Shafer structure on the real line $\mathbb{R}$ can be identified with a basic probability assignment, that is a map as follows:

$$m : 2^{\mathbb{R}} \to [0, 1] \tag{1}$$

where the probability mass is $m([\underline{x}_i, \overline{x}_i]) = p_i$ for each focal element $[\underline{x}_i, \overline{x}_i] \subseteq \mathbb{R}$ with $i = 1, \ldots, n$. $m(S)$ is equal 0 for the empty set $S = \emptyset$ and for $S \neq [\underline{x}_i, \overline{x}_i]$, such that $p_i > 0\,\forall i$ and $\sum_i p_i = 1$. The upper bound on probability is referred as plausibility and the lower bound as belief, the cumulative plausibility function $Pl(x)$ and cumulative belief function $Bel(x)$ can be computed as $Pl(x) = \sum_{\underline{x}_i \leq x} m_i$ and $Bel(x) = \sum_{\overline{x}_i \leq x} m_i$. The continuous equivalents of DS structures are the so-called probability boxes or P-boxes. Mathematically, a P-box is a pair of lower and upper cumulative distribution functions $[\underline{F}_X, \overline{F}_X]$ from the possibility space Θ to [0,1] such that $\underline{F}_X(x) \leq \overline{F}_X(x)\ \forall\ x \in \Omega$ and Ω is a classical probability space. The upper and lower bounds for the CDFs are $\overline{F}_X = \overline{P}(X \leq x)$ and $\underline{F}_X = \underline{P}(X \leq x)$, respectively. Note that the probability distribution family associated with the random variable x can be either specified or not specified. The former are generally named distributional P-boxes, or parametric P-boxes, the latter are named distribution-free P-boxes, or non-parametric P-boxes [13]. The wider the distance between the upper and the lower bound is, the higher the incertitude associated to the random variable. P-boxes and DS structures offer a straightforward way to deal with multiple and overlapping intervals, inconsistent sources of information and small sample sizes. The drawback is that the computational cost of propagating P-boxes and DS structures through the system is generally quite high, especially for a large number of intervals (i.e. focal elements) and time-consuming models. Nevertheless, the quantification approaches are generally not-intrusive and hence applicable to any model.