Distribution-free P-box processes: definition and simulation

Matthias G.R. Faes
KU Leuven, Department of Mechanical Engineering, LMSD Division, matthias.faes[at]kuleuven.be
Institute for Risk and Reliability, Leibniz Universität Hannover

Matteo Broggi
Institute for Risk and Reliability, Leibniz Universität Hannover, broggi[at]irz.uni-hannover.de

Kok-Kwang Phoon
National University of Singapore, Department of Civil & Environmental Engineering,kkphoon[at]nus.edu.sg

Michael Beer
Institute for Risk and Reliability, Leibniz Universität Hannover, beer[at]irz.uni-hannover.de
Institute for Risk and Uncertainty and School of Engineering, University of Liverpool, Peach Street, Liverpool L69 7ZF, UK
International Joint Research Center for Engineering Reliability and Stochastic Mechanics, Tongji University, 1239 Siping Road, Shanghai 200092, P.R. China

Abstract. Typically, non-deterministic models of spatial or time dependent uncertainty are modelled using the well-established random field framework. However, while tailored for exactly these types of time and spatial variations, stochastic processes and random fields currently have only limited success in industrial engineering practice. This is mainly caused by its computational burden, which renders the analysis of industrially sized problems very challenging, even when resorting to highly efficient random field analysis methods such as EOLE. Apart from that, also the methodological complexity, high information demand and rather indirect control of the spatial (or time) variation has limited its cost-benefit potential for potential end-users. This data requirement was recently relaxed by the first author with the introduction of imprecise random fields, but so far the method is only applicable to parametric p-box valued stochastic processes and random fields. This paper extends these concepts by expanding the framework towards distribution-free p-boxes. The main challenges addressed in this contribution are related to both the non-Gaussianity of realisations of the imprecise random field in between the p-box bounds, as well as maintaining the imposed auto-correlation structure while sampling from the p-box. A case study involving a dynamical model of a car suspension is included to illustrate the presented concepts.

Keywords: uncertainty quantification, stochastic processes, imprecise probabilities

1. Introduction

1.1. General rationale

Stochastic processes are widely used in domains such as engineering (Stefanou, 2009) and financial economics to represent stochastic quantities that vary over time- and/or space (Vanmarcke, 1983). However, due to the theoretical and computational difficulties, usually these processes are assumed to be Gaussian, which might not always be a truthful representation of reality (Grigoriu, 2010). Furthermore, the definition of a stochastic process requires the rigorous description of the governing distribution function, which includes selecting the appropriate distribution family as well as the governing hyper-parameters. In practice, this might not always be possible due to limitations on the available data (quantity of the data, corrupted or missing data, etc.), but also conflicting sources of information (e.g., expert opinions). Recently introduced approaches based on Bayesian compressed sensing alleviate this problem (see e.g., (Wang et al., 2018; Montoya-Noguera et al., 2018) by making the estimated stochastic process robust to missing data, even when these processes are non-Gaussian and have an unknown non-stationary auto-correlation function (Wang et al., 2019). Furthermore, also detrending is not required in these cases (Wang et al., 2019). In (Ching and Phoon, 2020), these methods were extended to also account for the general case of multivariate, uncertain, unique, sparse, incomplete spatial data (denoted MUSIC-X by the authors).

As a possible alternative pathway to the issue of low data availability is to resort to the more general framework of imprecise probabilities (Beer, Ferson and Kreinovich , 2013). According to this framework, epistemic uncertainty resulting from potential data deficiencies are taken explicitly into account to allow for rigorous analysis. In the context of processes, parametric p-box valued stochastic processes have been introduced in e.g., (Dannert et al., 2018), (Faes and Moens, 2019) and (Fina et al., 2020). However, such approaches still require the definition of a distribution family, which is not always possible. Furthermore, these approaches are based on Gaussian distributed processes. This paper aims to go further than the available methods for simulating from imprecise stochastic processes by introducing a type of distribution-free p-box stochastic process. Such process is obtained by passing a standard normal Gaussian stochastic process through an interval-valued translation map, effectively providing a distribution-free p-box stochastic process. An example is included to illustrate the definition and propagation of these structures. The paper is constructed as follows: the remainder of this section recalls some important concepts concerning the definition and simulation of non-Gaussian stochastic processes; Section 2 introduces the approach for defining distribution-free p-box processes; Section 3 briefly discusses a double-loop approach to propagate these structures; Section 4 provides a case study as illustration of the approach; Section 5 lists the conclusions of the work.

1.2. Stochastic processes

A finite-dimensional stochastic process $x(t,\omega)$ describes a set of correlated random variables $x(\omega)$, which are assigned to a number of locations $t \in \Omega_d$ in the model domain $\Omega_d \subset \mathbb{R}^d$ with dimension $d \in \mathbb{N}$. Note that Ω_d may comprise both space and/or time dimensions. Each random variable $x(\omega) : (\Omega, \varsigma, P) \mapsto \mathbb{R}$ as such maps from a complete probability space to the real domain, with $\omega \in \Omega$ a coordinate in sample space Ω and ς the sigma-algebra. In this paper, we require that $x(t,\Omega) \in$

$\mathcal{L}^2(\Omega, P)$, with $\mathcal{L}^2(\Omega, P)$ the Hilbert space of second-order random variables (i.e., finite variance). For a given event $\omega_i \in \Omega$, the corresponding $x(t, \omega_i)$ is a realization of the stochastic process. A stochastic process is considered Gaussian if the distribution of $(x(t_1, \omega), x(t_2, \omega), \ldots, x(t_{n_t}, \omega))$, with $n_t \in \mathbb{N}$, is jointly Gaussian $\forall t \in \Omega_d$. In this case, $x(t, \omega)$ is completely described by its mean function $\mu_x(t) : \Omega_d \mapsto \mathbb{R} = E_\omega[x(t, \omega)]$ and its auto-covariance function $\mathbf{C}_{xx}(t, t') : \Omega_d \times \Omega_d \mapsto \mathbb{R}$, given by $\mathbf{C}_{xx}(t, t') = E_\omega[x(t, \omega), x(t', \omega)]$ (Vanmarcke, 1983). In the remainder of the paper, univariate stochastic processes (i.e., $\Omega_d \in \mathbb{R}^1$) are considered for the sake of conciseness of notation. Note that the concepts explained in this paper scale straightforwardly to multivariate or multi-index stochastic processes (e.g., random fields) as well.

Generally, when applying stochastic fields in an engineering context, for instance to represent a spatially uncertain input quantity of a finite element model, the field has to be discretized over Ω_d. In this context, the Karhunen-Loève expansion is a very powerful tool to represent stochastic fields (Spanos, 1989). Specifically, following the Karhunen-Loève (KL) series expansion, a stochastic process $x(t, \omega)$ is represented as:

$$x(t, \omega) = \mu_x(t) + \sigma_x \sum_{i=1}^{\infty} \sqrt{\lambda_i} \psi_i(t) \xi_i(\omega), \tag{1}$$

with σ_x the standard deviation of the random field and where the quantities $\lambda_i \in (0, \infty)$ and $\psi_i(t) : \Omega_d \mapsto \mathbb{R}$ are respectively the eigenvalues and eigenfunctions of the continuous, bounded, symmetric and positive (semi-)definite stationary auto-correlation function $\rho_{xx}(\tau) : \Omega_d \times \Omega_d \mapsto [0, 1]$, where Ω_d is a bounded and closed interval, in accordance with Mercer's theorem:

$$\rho_{xx}(\tau) = \sum_{i=1}^{\infty} \lambda_i \psi_i(t) \psi_i(t'), \tag{2}$$

These quantities are in practice obtained by solving the homogeneous Fredholm integral equation of the second kind:

$$\int_{\Omega_d} \rho_{xx}(\tau) \psi_i(t') dt' = \lambda_i \psi_i(t), \tag{3}$$

where $t' = t + \tau$ for which many efficient discretization schemes exist (Betz et al., 2014). Since $\rho_{xx}(\tau)$ is bounded, symmetric and positive semi-definite, and furthermore in most practical cases can be assumed positive definite, these eigenvalues λ_i are non-negative and the eigenfunctions $\psi_i(t)$ satisfy the following orthogonality condition:

$$\langle \psi_i(t), \psi_j(t) \rangle = \int_{\Omega_d} \psi_i(t) \psi_j(t) dt = \delta_{ij} \tag{4}$$

with δ_{ij} the Kronecker delta and $\langle \cdot, \cdot \rangle : \Omega_d \times \Omega_d \mapsto \mathbb{R}$ an inner product on the space $\mathcal{L}^2(\Omega, dt)$. In this case, the series expansion in Eq. 2 can be shown to be optimally convergent (Spanos, 1989).

The variables $\xi_i(\omega), i = 1, \ldots, \infty$, introduced in Eq. 1, are uncorrelated random variables, which are determined according to:

$$\xi_i(\omega) = \frac{1}{\sqrt{\lambda_i}} \int_{\Omega_d} [x(t, \omega) - \mu_x(t)] \psi_i(t) dt, \tag{5}$$

which can be shown to be independent standard normally distributed in the case of a Gaussian random field. For practical reasons, the infinite series expansion in Eq. 1 is usually truncated after a finite number of terms $n_{KL} \in \mathbb{N}$:

$$x(t,\omega) = \mu_x(t) + \sigma_x \sum_{i=1}^{n_{KL}} \sqrt{\lambda_i}\psi_i(t)\xi_i(\omega), \tag{6}$$

where n_{KL} should be selected such that a well-chosen variance error metric is minimized.

For non-Gaussian fields, the $\xi_i(\omega), i = 1, \ldots, n_{KL}$ represented in Eq. 5 are non-Gaussian too and their distribution needs to be solved for explicitly. Furthermore, in this case, the corresponding random variables $\xi_i(\omega)$ may be uncorrelated but dependent. These dependencies may be difficult to quantify (Grigoriu, 2010). Finally, Eq. 5 reveals that the distribution of $\xi_i(\omega)$ depends on sample path realisations $x(t,\omega)$ of the stochastic process, and hence, iterative methods such as presented by (Huang, Quek and Phoon, 2001) need to be applied. Alternatively, also translation theory as introduced by Grigoriu (Grigoriu, 1998) provides a viable approach towards simulating from (strongly) non-Gaussian stochastic processes.

1.3. Translation stochastic fields

Translation process theory, as introduced by Grigoriu (Grigoriu, 1998), provides a different pathway for the simulation of non-Gaussian stochastic fields. Specifically, a stationary Gaussian stochastic process $\eta(t,\omega)$ with autocorrelation function $\rho_{xx}(\tau)$ is transformed using a nonlinear transformation into a non-Gaussian field $x(t,\omega)$, which is formally expressed as:

$$x(t,\omega) = F_X^{-1} \circ \Phi(\eta(t,\omega)) = g(\eta(t,\omega)) \tag{7}$$

with $g := F_X^{-1} \circ \Phi$ the so-called translation mapping, F_X^{-1} the inverse of the target non-Gaussian cumulative distribution function (CDF) that represents the distribution of the non-Gaussian stochastic process and Φ the marginal standard normal CDF, i.e., $\Phi(\eta) = P(\eta(t,\omega) < \eta)$.

It can be shown (see (Grigoriu, 1998)) that the mean μ_x, variance σ_x^2 and correlation function $r_{xx}(\tau)$ of $x(t,\omega)$ have closed-form solutions, that are given respectively by:

$$\mu_x = E_\eta[x(t,\omega)] = E_\eta[g(\eta(t,\omega))] = \int_{-\infty}^{\infty} g(\eta)\phi(\eta)d\eta \tag{8}$$

$$\sigma_x^2 = E_\eta[x(t,\omega) - \mu_x]^2 = \int_{-\infty}^{\infty} (g(\eta) - \mu_x)^2\phi(\eta)d\eta \tag{9}$$

$$r_{xx}(\tau) = E[x(t,\omega) - \mu_x]E[x(t+\tau,\omega) - \mu_x] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (g(\eta_1) - \mu_x)(g(\eta_2) - \mu_x)\phi(\eta_1, \eta_2, \rho_{xx}(\tau))dydz \tag{10}$$

with $\phi(\eta_1, \eta_2, \rho_{xx}(\tau))$ the density of a bivariate standard Gaussian vector with correlation coefficient $\rho_{xx}(\tau)$, given by:

$$\phi(\eta_1, \eta_2, \rho_{xx}(\tau)) = \frac{1}{\sqrt{2\pi(1-\rho_{xx}(\tau)^2)}} \exp -\frac{\eta_1^2 + \eta_2^2 - 2\eta_1\eta_2\rho_{xx}(\tau)}{2(1-\rho_{xx}(\tau))^2} \tag{11}$$

These equations can be readily solved by standard numerical quadrature schemes. Furthermore, the forward transformation of a Gaussian process to a non-Gaussian process is always possible (Grigoriu, 1998).

On the other hand, given a non-Gaussian $r_{xx}(\tau)$, it is not always possible to determine the corresponding Gaussian autocorrelation $\rho_{xx}(\tau)$ that, when transformed, yields $r_{xx}(\tau)$ (Grigoriu, 1998). This happens when either the inverse of Eq. 10 yields an autocorrelation function that is not positive semi-definite, or when the normalized autocorrelation $\xi(\tau)$ has values that lie outside of the admissible range $[\xi(\tau^{min}), \xi(\tau^{max})]$ which can be found by setting $\rho_{xx}(\tau)$ to respectively 1 and -1 in Eq. 10. In literature, approximative methods have been introduced to find a representative $r_{xx}(\tau)$ that is an admissible autocorrelation function (see e.g., (Kim and Shields, 2015)).

2. P-box stochastic fields

In practical engineering cases, it is not always possible to define a crisp distribution function $F_X(x)$ to construct the non-Gaussian stochastic process. In this context, p-boxes can provide a valuable tool to represent the uncertainty an analysist has on the specification of the appropriate distribution function. This section introduces p-box-valued stochastic processes.

A scalar distribution-free p-box is usually described by a lower CDF $\underline{F}_X \in \mathbb{F}$ and an upper CDF $\overline{F}_X \in \mathbb{F}$, where $\mathbb{F}$ expresses the the set of all CDFs on $\mathbb{R}$. They are collected as a pair $\left[\underline{F}_X, \overline{F}_X\right]$ which yields a set of possible CDFs via $\underline{F}_X(x) \leq F_X(x) \leq \overline{F}_X(x),\ x \in \mathbb{R}$. A distribution-free p-box as such corresponds to defining a lower probability $\underline{P}$ and upper probability $\overline{P}$ on events $\{X \leq x\} = (-\infty, x]$, i.e., $\underline{P}(X \leq x) = \underline{F}_X(x)$ and $\overline{P}(X \leq x) = \overline{F}_X(x)$ for $x \in \mathbb{R}$, which define a credal set of probability measures. In case additional information on the uncertainty is available, constraints on the p-box can be enforced. For instance, if the (class of) distribution functions $\mathcal{F}$ is known, the set of possible CDFs $\{F_X(\cdot, \theta) \in \mathcal{F} \mid \theta \in D_\theta\}$ can be defined conditional on a vector of hyper-parameters θ. Since this is a special case of the distribution-free p-box, the following discussion on distribution-free p-box stochastic fields is equally applicable.

As explained in Section 1, a (precise) stochastic process $x(t, \omega)$ can be considered as a collection of n_t correlated random variables distributed throughout the model domain Ω_d. Imprecise stochastic processes and fields $\hat{x}(t, \omega)$ can be regarded as a natural extension of this idea, where for each discrete location $t_i \in \Omega_d$, a scalar p-box is defined. However, since all $x(t_i, \omega)$ are correlated according to $\rho_x(\tau)$, these scalar p-boxes also are correlated to each other. This observation complicates the analysis of distribution-free imprecise stochastic processes drastically, since the direct simulation from a set of correlated distribution-free p-boxes is far from trivial from both a theoretical as a numerical point of view. Furthermore, since distribution-free p-boxes are considered, also non-Gaussian processes are inherently included in the imprecise stochastic description.

A potential solution to this issue is to start from a precise standard normal Gaussian stochastic process $\eta(t, \omega)$ with predefined correlation function $\rho_{xx}(\tau)$ and pass this representation through an imprecisely defined translation map which is defined as:

$$\hat{x}(t, \omega) = \left[\underline{F}_X^{-1}, \overline{F}_X^{-1}\right] \circ \Phi(\eta(t, \omega)), \tag{12}$$

which can be further expanded as:

$$\hat{x}(t,\omega) = \left[\underline{F}_X^{-1}, \overline{F}_X^{-1}\right] \circ \Phi\left(\sum_{i=1}^{n_{KL}} \sqrt{\lambda_i}\psi_i(t)\xi_i(\omega)\right), \tag{13}$$

where $\Phi\left(\sum_{i=1}^{n_{KL}} \sqrt{\lambda_i}\psi_i(t)\xi_i(\omega)\right)$ represents a re-scaling of $\eta(t,\omega)$ to the interval $[0,1]$. Since a CDF is by definition a monotonic function, i.e., $F_X(x_1) \leq F_X(x_2) \iff x_1 < x_2$, the bounds of $\hat{x}(t,\omega)$ are determined by $\left[\underline{F}_X^{-1}, \overline{F}_X^{-1}\right]$. Note that in case F_X is not strictly monotonic (i.e.,$F_X(x_1) < F_X(x_2) \iff x_1 < x_2$, pseudo-inverses should be used to calculate the inverse of F_X. Furthermore, since $\left[\underline{F}_X^{-1}, \overline{F}_X^{-1}\right] \circ \Phi$ represents an interval-valued mapping, each realisation of the stochastic process $\eta(t,\omega_j)$, corresponding to the event ω_j is translated towards an interval field that is consistent with the bounds on the CDF (due to the monotonicity of the CDF), and which is given as:

$$x^I(t,\omega_j) = \left[\underline{F}_X^{-1}, \overline{F}_X^{-1}\right] \circ \Phi\left(\sum_{i=1}^{n_{KL}} \sqrt{\lambda_i}\psi_i(t)\xi_i(\omega_j)\right), \tag{14}$$

with the lower bound given as:

$$\underline{x}(t,\omega_j) = \overline{F}_X^{-1} \circ \Phi\left(\sum_{i=1}^{n_{KL}} \sqrt{\lambda_i}\psi_i(t)\xi_i(\omega_j)\right), \tag{15}$$

and the upper bound defined as:

$$\overline{x}(t,\omega_j) = \underline{F}_X^{-1} \circ \Phi\left(\sum_{i=1}^{n_{KL}} \sqrt{\lambda_i}\psi_i(t)\xi_i(\omega_j)\right). \tag{16}$$

It should be noted that this is not an explicit interval field, i.e., an interval field that is represented as a series expansion with interval-valued coefficients as described in (Faes and Moens, 2020a), since the interval-valued nature in this field stems from the mapping that is performed on a single realisation of the crisp Gaussian field, rather than from a series expansion with interval-valued weights. As such, typically applied interval propagation methods, as described in (Faes and Moens, 2020b), cannot be applied straightforwardly to propagate this interval field. Furthermore, the auto-dependence function of realisations within this interval field becomes interval-valued too. Note that in case of intervals, auto-dependence is used to describe the dependency throughout Ω_d, rather than autocorrelation or autocovariance. Intermediate realisations of $x^k(t,\omega_j) \in x^I(t,\omega_j)$ can be generated by drawing admissible CDFs $F_X^{-1,k} \in \left[\underline{F}_X^{-1}, \overline{F}_X^{-1}\right]$. For reasons of clarity, the explanation of a possible procedure to do so is deferred to a later section.

Conversely, when collecting all $x^k(t,\omega)$, that correspond to a certain realisation $F_X^{-1,k} \in [\underline{F}_X^{-1}, \overline{F}_X^{-1}]$, this becomes again a crisp random field, the properties of which can be computed by virtue of translation field theory as:

$$\mu_{x^k(t,\omega)} = \int_{-\infty}^{\infty} F_X^{-1,k} \circ \Phi(\eta)\phi(\eta)d\eta \tag{17}$$

$$\sigma^2_{x^k(t,\omega)} = \int_{-\infty}^{\infty} (F_X^{-1,k} \circ \Phi(\eta) - \mu_{x^k(t,\omega)})^2 \phi(\eta) d\eta \tag{18}$$

$$r_{x^k x^k}(\tau) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (F_X^{-1,k} \circ \Phi(\eta_1) - \mu_{x^k(t,\omega)})(F_X^{-1,k} \circ \Phi(\eta_2) - \mu_{x^k(t',\omega)}) \phi(\eta_1, \eta_2, \rho_{xx}(\tau)) d\eta_1 d\eta_2 \tag{19}$$

with $\phi(\eta_1, \eta_2, \rho_{xx}(\tau))$ as defined in Eq. (11). Note that this random field is stationary in case the underlying Gaussian field is stationary since it is mapped through a crisp $F_X^{-1,k}$. Evidently every $F_X^{-1,k}$ will yield a random field with generally different central moments for every $F_X^{-1,k} \in [\underline{F}_X^{-1}, \overline{F}_X^{-1}]$.

Based on the preceding discussion, it can be seen that an imprecise stochastic process can jointly be regarded as a stochastic collection of interval fields as well as as a credal set of stochastic fields. Since each stochastic realisation of the imprecise stochastic process is an interval field, and vice versa, each realisation within the p-box corresponds to a stochastic process, also the mean of the corresponding imprecise field is interval-valued. Due to the monotonicity of the translation map, this mean is formally given as:

$$\mu_x^I = [\underline{\mu}_x, \overline{\mu}_x] = \left[E_\omega\left[\underline{x}(t,\omega)\right], E_\omega\left[\overline{x}(t,\omega)\right]\right]. \tag{20}$$

These bounds can be computed in a straightforward manner by invoking Eq. 17 twice: once on the ensemble of lower bounds and once on the ensemble of upper bounds. As an illustration of these concepts, consider the example in Figure 1. This figure shows two stochastic realisations of the crisp zero-mean Gaussian stochastic process, as well as their transformation into two realisations of the distribution-free p-box random field, which manifest themselves in the shape of two interval fields. From this figure, it can also be seen that the collection of all upper bounds of these interval fields, denoted $\overline{x}(t,\omega)$, tends towards a stochastic process that has $\underline{F}_X(x)$ as a distribution when the number of terms in the KL expansion tends to infinity. The same obviously holds for the lower bound, as well as any intermediate realisation $F_X^{-1,k} \in \left[\underline{F}_X^{-1}, \overline{F}_X^{-1}\right]$.

As a final comment, it should be noted that the auto-correlation function of $x^I(t,\omega)$, being $r_{xx}(\tau)$ also has become an interval field due to the interval-valued translation map. This might have important implications for structural dynamical problems, where the match of a dominant frequency of the loading process might interfere with a natural frequency of the structure. The detailed treatment of this issue and a potential solution hereto however fall outside the scope of this paper.

3. Propagation of distribution-free imprecise stochastic fields

Usually, an analyst who is confronted with imprecise probabilistic model quantities is concerned with finding the bounds on some probabilistic measure $\mathcal{P}$ of the model's responses of interest. In case a crisp density function $f_{\mathbf{X}}$ is known, the n^{th} central moment of the model's response $E_{\mathbf{y}}[Y^n]$ or the probability of failure p_F is determined by evaluating an integral equation of the following form:

$$\mathcal{P} = \int_{\mathbb{R}^{n_{KL}}} \mathcal{H}(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) \, \mathrm{d}\mathbf{x}, \tag{21}$$

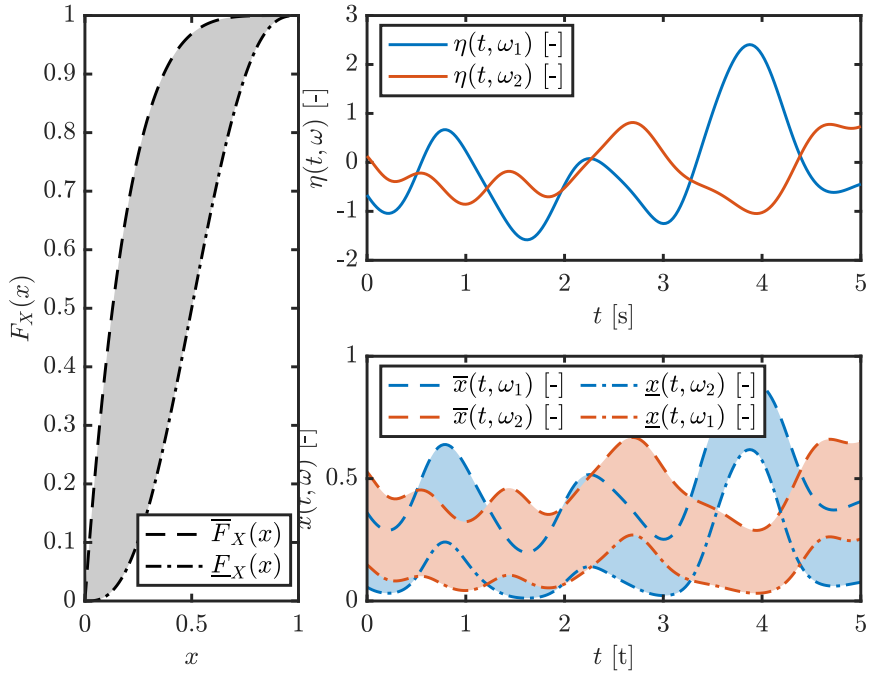

Figure 1. Illustration of the transformation of two stochastic realisations of the crisp zero-mean Gaussian stochastic process into two realisations of the distribution-free p-box random field, which manifest themselves in the shape of two interval fields.

where $\mathcal{P}$ denotes, depending on the context, the n^{th} central moment of the model's response $E_{\mathbf{y}}[Y^n]$ or the probability of failure p_F. In this equation, the random variable $\mathbf{X} = [\xi_1, \xi_2, \ldots, \xi_{n_{KL}}]$ represents a vector of i.i.d. standard normal random variables. These stem from the finite-dimensional approximation of the random field $\eta(t,\omega)$, as given in Eq. (6). Hence, $f_{\mathbf{X}}(\mathbf{x}) = \prod_{i=1}^{n_{KL}} \phi(x_i)$. In case $\mathcal{P} \equiv E_{\mathbf{y}}[Y^n]$ is considered, $\mathcal{H} \equiv r^n(\mathbf{x})$, where $r(\mathbf{x})$ represents the so-called performance function of the model. On the other hand, in case the calculation is aimed at computing p_F, $\mathcal{H} \equiv I_r(\mathbf{x})$, where I_r is the indicator function which is 1 in case $r(\mathbf{x}) \leq 0,\ \mathbf{x} \in \mathbb{R}^{n_x}$, and 0 otherwise. In this context, $\mathbf{x}$ is used to denote

In the specific case of distribution-free imprecise random fields, the probabilistic measure $\mathcal{P}$ becomes conditional on the realisation F_X^k of the p-box that represents the translation map CDFs, i.e., $\mathcal{P}(F_X^k)$. As explained in Section 2, the distribution of a random field realisation $x^k(t,\omega)$ of the imprecise random field $\hat{x}(t,\omega)$, as described in Eq. (13) tends towards F_X^k as $n_{KL} \to \infty$. In this case, Eq. (22) reads:

$$\mathcal{P}(F_X^k) = \int_{\mathbb{R}^{n_{KL}}} \mathcal{H}\left(F^{-1,k} \circ \Phi\left(\boldsymbol{\Psi}\sqrt{\boldsymbol{\Lambda}}\mathbf{x}\right)\right) f_{\mathbf{X}}^k(\mathbf{x})\,\mathrm{d}\mathbf{x}, \tag{22}$$

where f_X^k represents the i.i.d. standard normal random variables that have been passed through the translation map and $\boldsymbol{\Psi}$ and $\boldsymbol{\Lambda}$ representing respectively a matrix and vector collecting the

eigenfunctions and -values of the autocorrelation function of the underlying Gaussian field. The main point in calculating this time or space-dependent probabilistic measure lies in the definition of the performance function. For instance, the performance function can be defined such that it converts the time-dependent problem into a series event by considering a first passage probability.

To infer the bounds on $\mathcal{P}$, two optimization problems need to be solved to actively search the parameter space spanned by $[\underline{F}_X(x), \overline{F}_X(x)]$. Specifically, the lower bound is obtained as:

$$\underline{\mathcal{P}} = \min \mathcal{P}(F_X^k), \tag{23}$$

where the minimum is taken over all distribution functions F_X^k such that $[F_X^k \in \underline{F}_X(x), \overline{F}_X(x)]\ \forall x$. Sinilary, the upper bound is determined as:

$$\overline{\mathcal{P}} = \max \mathcal{P}(F_X^k) \tag{24}$$

Note that each realisation F_X^k drawn from this interval represents a non-Gaussian random field with auto-correlation structure as described in Eq. (10). As such, this is effectively a double-loop approach, which might entail a non-negligible computational cost to solve. The main difficulty associated with solving these optimization problems lies in the fact that the optimisation has to be performed over the infinite-dimensional space of bounded, strictly monotonically increasing functions over the support of x. Such calculation is intractable, even for the most simple cases. In this paper, it is therefore proposed to approximate these optimization problems as discrete problems. Specifically, it is aimed at solving following problems:

$$\begin{aligned} \underline{\mathcal{P}} &= \min_{\mathbf{F}_X \in \mathbf{F}_X^I} \mathcal{P}(F_X^k) \\ \overline{\mathcal{P}} &= \max_{\mathbf{F}_X \in \mathbf{F}_X^I} \mathcal{P}(F_X^k) \end{aligned} \tag{25}$$

subject to:

$$\mathbf{A}\mathbf{F}_X \leq 0 \tag{26}$$

where $\mathbf{F}_X = F_X(\mathbf{x}_s)$, with $\mathbf{x}_s \in \mathbb{R}^{n_s}$ representing n_s equally spaced sample points throughout the support of $[\underline{F}_X(x), \overline{F}_X(x)]$. Similarly, $\mathbf{F}_X^I = [\underline{F}_X(\mathbf{x}_s), \overline{F}_X(\mathbf{x}_s)] \in \mathbb{IR}^{n_s}$ represents an n_s dimensional interval vector collecting the bounds of the p-box for each sample point in the support. The inequality shown in Eq. (26) enforces the realisations drawn from the interval vector $\mathbf{F}_X^I$ to be strictly monotonic to ensure that they represent admissible CDFs, where $\mathbf{A} \in \mathbb{R}^{n_s-1 \times n_s}$ represents an upper-triangular band matrix with $A_{1,:} = [1 \ -1 \ 0 \ \dots \ 0]$. As such, the infinite-dimensional optimisation problem is converted to a linear-inequality-constrained optimization problem over n_s variables. Finally, it can be noted that the translation mapping explained in section 2 requires the calculation of the inverse of the CDF (see e.g., Eq. 17). This is for instance required to generate the required sample paths of $x^k(t,\omega)$ to estimate the performance function $r(\mathbf{x}^k(t,\omega))$. Hereto, a piece-wise cubic Hermite polynomial interpolation is performed using the Fritsch-Carlson algorithm to estimate a functional relationship between $\mathbf{x}_s$ and $\mathbf{F}_X$. This approach is selected as this allows for generating a strictly monotonic, C^1 continuous interpolation of the inverse of $\mathbf{F}_X(\mathbf{x}_s)$ (Fritsch and Carlson, 1980).

4. Case study

The case study represents a quarter-car model, which is a 2-DOF idealisation of the realistic dynamics of the suspension of a car. Specifically, this case study is concerned with assessing the bounds on several comfort metrics of a vehicle suspension, given a p-box process-valued base excitation. The quarter-car dynamics can be represented as a set of two ordinary differential equations:

$$m_s\ddot{x}_s + c_s(\dot{x}_s - \dot{x}_{us}) + k_s(x_s - x_{us}) = 0 \tag{27}$$

$$m_{us}\ddot{x}_{us} - c_s(\dot{x}_s - \dot{x}_{us}) - k_s(x_s - x_{us}) + c_t(\dot{x}_{us} - \dot{x}_0) + k_t(x_{us} - x_0) = 0 \tag{28}$$

with $\dot{\bullet}$ the time derivative of $\bullet$, x_{us} the displacement of the unsprung mass (i.e., the suspension components, wheel and other components directly connected to them), x_s the displacement of the sprung mass (i.e., all components resting on the suspension), m_{us} and m_s the unsprung and sprung mass of a quarter of the car, c_s and c_t respectively the damping coefficients of the suspension and tire, k_s and k_t respectively the stiffness coefficients of the suspension and tire. Finally, x_0 and $\dot{x}_0$ are the displacement and velocity in vertical direction that excite the bottom of the wheel (i.e., the road profile). The complete road profile is denoted $x_0(t)$. The dynamics of the car are simulated over a distance of 50 (m), when the car is travelling at a speed of 10 (m/s). The one-dimensional spatial domain is discretized into 1000 equidistant points and the time domain is discretized into time intervals of 0.005 (s). A schematic representation of the model is given in figure 2.

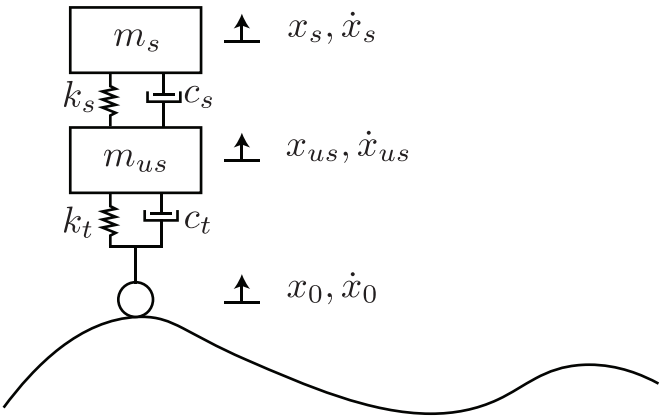

Figure 2. Schematic illustration of the quarter-car model

For the solution of this coupled system of ODEs, a state-space model is employed:

$$\frac{d}{dt}\begin{bmatrix} x_{us} - x_0 \\ \dot{x}_{us} \\ x_s - x_{us} \\ \dot{x}_s \end{bmatrix} = A\begin{bmatrix} x_{us} - x_0 \\ \dot{x}_{us} \\ x_s - x_{us} \\ \dot{x}_s \end{bmatrix} + \begin{bmatrix} -1 \\ \frac{c_t}{m_{us}} \\ 0 \\ 0 \end{bmatrix}\dot{x}_0 \tag{29}$$

with the matrix A equal to:

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ \frac{-k_t}{m_{us}} & \frac{-(c_s+c_t)}{m_{us}} & \frac{k_s}{m_{us}} & \frac{c_s}{m_{us}} \\ 0 & -1 & 0 & 1 \\ 0 & \frac{c_s}{m_s} & \frac{-k_s}{m_s} & \frac{-c_s}{m_s} \end{bmatrix} \tag{30}$$

Four state variables are considered, being respectively the tire deflection $(x_{us} - x_0)$; the unsprung mass velocity $\dot{x}_{us}$; the suspension stroke $x_s - x_{us}$, and sprung mass velocity $\dot{x}_s$. Typically, in the context of assessing the dynamical comfort of a car, two parameters are of interest: the suspension stroke (i.e., the relative displacement of the car body with respect to the unsprung mass) and the acceleration of the sprung mass. In the proceeding study, the damping effect of the tire, c_t is considered negligible.

Table I. Considered case studies for the distribution free p-box field

Case	$\underline{F}_X(x)$	$\overline{F}_X(x)$	b_t
1	$\min[\mathcal{B}(1,3), \mathcal{B}(5,5)])$	$\max[\mathcal{B}(1,3), \mathcal{B}(5,5)])$	0.5
2	$\min[\mathcal{B}(3,1), \mathcal{B}(5,5)])$	$\max[\mathcal{B}(3,1), \mathcal{B}(5,5)])$	0.5
3	$\min[\mathcal{B}(1,1), \mathcal{B}(2,5)])$	$\max[\mathcal{B}(1,1), \mathcal{B}(2,5)])$	0.5
4	$\min[\mathcal{B}(1,0.2), \mathcal{B}(5,5)])$	$\max[\mathcal{B}(1,.02), \mathcal{B}(5,5)])$	0.5
5	$\min[\mathcal{N}(0,0.75), \mathcal{B}(1,0.2)])$	$\max[\mathcal{N}(0,0.75), \mathcal{B}(1,0.2)])$	2

The complete road profile $x_0(t)$ is modelled as a p-box valued stochastic process. The auto-correlation of the underlying Gaussian process $\eta(t,\omega)$ is governed by a squared exponential auto-correlation function with a correlation length of 0.5 m. Sample path realisations of $\eta(t,\omega)$ are generated using the Karhunen–Loève series expansion, while retaining 32 terms. The stochastic content of the imprecise stochastic process is represented using a distribution-free p-box. For illustrative reasons, the bounds of the p-box, $[\underline{F}_X(x), \overline{F}_X(x)]$ are generated by taking the extremes of a set of distributions. Hereto, 5 different case studies are considered, which are summarized in Table I. The corresponding p-boxes area also visualized in figure 3. Note that these cases do not necessarily represent a physical phenomenon. Rather, they are selected for illustrative reasons. Each random field corresponding to a realisations of these p-boxes is given as:

$$x_0^k(t,\omega) = (F_X^k)^{-1} \circ \Phi\left(\sum_{i=1}^{n_{KL}} \sqrt{\lambda_i}\psi_i(t)\xi_i(\omega)\right) \tag{31}$$

which are generated by the optimization algorithms introduced in Section 4. Applying the double-loop optimization algorithm introduced in Section 4, the bounds on the probability of failure of the structure are computed. In this context, the performance function $r(\mathbf{x})$ of the car model is given as:

$$r(x_0) = 1 - \max_{i=1,\ldots,m}\left(\frac{|x_s(\mathbf{x},t_i) - x_{us}(\mathbf{x},t_i)|}{b_t}\right) \tag{32}$$

where the threshold value b_t is also given in Table I. This corresponds to a first passage probability. Since the process is non-Gaussian, highly efficient and dedicated sampling methods such as Directional Importance Sampling (Misraji et al., 2020), as also applied in the context of imprecise probabilities in (Faes et al., 2020a) or (Faes et al., 2020b), are not applicable. Therefore, the integral equation in the inner loop of the optimization is solved using Subset-∞, as presented in (Au & Patelli, 2016), with an initial sample size of 5000 and a proposal standard deviation of 0.1. The discretisation of $F_X(\mathbf{x}_s)$, as described in Section 4, is performed using $n_s = 40$ slices, yielding a 40-dimensional optimization problem, which is solved using a gradient-free pattern search optimization algorithm. Pattern search is specifically selected to avoid the need to calculate gradients of p_F.

The results of performing the double-loop optimization problem are shown in Table II. In this table, $\underline{p}_F^*$ and $\overline{p}_F^*$ indicate the bounds on p_F obtained by means of optimization, whereas $\underline{p}_F$ and $\overline{p}_F$ are the failure probabilities corresponding to the bounds of the p-box. As is clear, the bounds obtained by just propagating $\underline{F}_X(x)$ and $\overline{F}_X(x)$ are not conservative. This is a direct result from the fact that the car model acts as a filter on the excitation towards the responses of interest.

Table II. Bounds on the probability of failure based on propagating the bounds, as well as performing optimization.

Case	$\underline{p}_F$	$\overline{p}_F$	$\underline{p}_F^*$	$\overline{p}_F^*$
1	$3.25 \cdot 10^{-4}$	0.065	$2.77 \cdot 10^{-4}$	0.385
2	$3.25 \cdot 10^{-4}$	0.061	$9.95 \cdot 10^{-6}$	0.378
3	0.0045	0.400	0.0028	0.624
4	0.0038	0.469	$3.01 \cdot 10^{-4}$	0.571
5	0.0062	0.015	$9.43 \cdot 10^{-6}$	0.423

A further explanation of these results can be given based on Figure 3. This figure shows clearly that the CDF corresponding to the highest probability of failure pushes the probability mass as much as possible towards the bounds of the p-box. This makes sense from a physical standpoint since the performance function contains an absolute value operation, and hence, positive and negative responses contribute both equally to the failure. Furthermore, the considered system is a 2-degree-of-freedom oscillator where the quantity of interest is the relative displacement between the two masses. This as such constitutes a perfect symmetric system. The CDF that minimizes p_F on the other hand aims at getting as much of the probability mass as possible towards the centre of the support.

5. Conclusions

This paper discusses the concept of distribution-free p-box stochastic processes and fields. To generate realisations of such a process, it is proposed to pass realisations of a standard Gaussian process through an imprecisely defined translation map such that the auto-correlation of the original

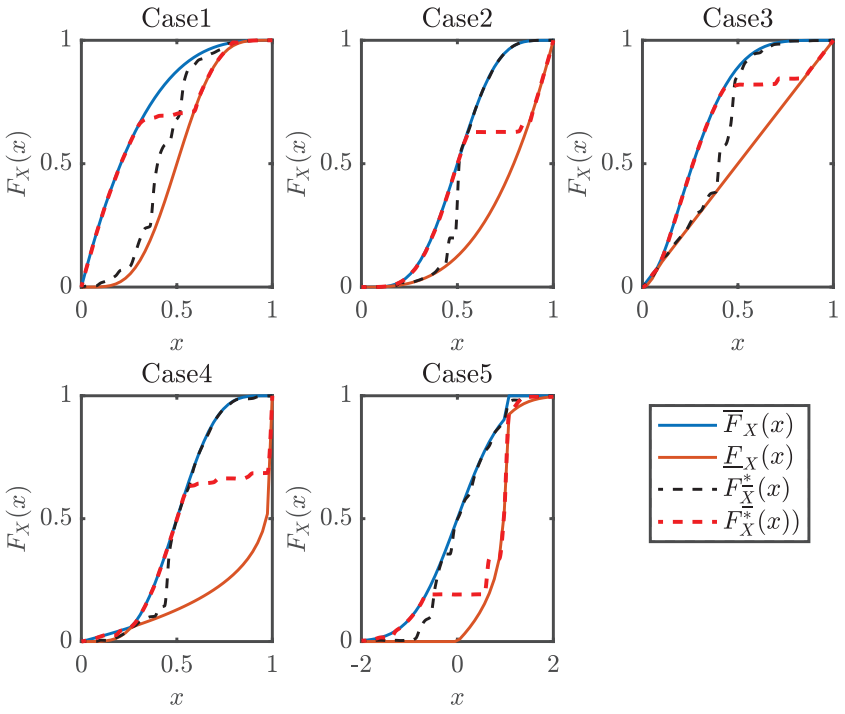

Figure 3. P-boxes corresponding to the 5 cases, as well as the realisations $F^*_{\underline{X}}$ and $F^*_{\overline{X}}$ that yield respectively $\underline{p}^*_F$ and $\overline{p}^*_F$.

process is largely maintained. Furthermore, an optimization approach is introduced to actively look for those realisations inside the p-box that yield a stochastic process that yields an extreme in a probabilistic measure of a response of interest. A case study on a quarter car model illustrated that the bounds of the P-box in fact do not necessarily coincide with the bounds on the probability of failure, which motivates the application of optimization algorithms.

Future work will focus on propagating the p-box field such that a target auto-correlation of the p-box field can be predefined by looking for an appropriate pre-mapped auto-correlation function for each realisation F^k_x in the p-box. Furthermore, the application of more advanced approaches such as e.g., based on sparse polynomial chaos expansions as discussed in (Schöbi and Sudret, 2017) to propagate the p-box will be investigated.

Acknowledgements

Matthias Faes acknowledges the financial support of the Research Foundation Flanders (FWO) under grant number 12P3519N as well as the Alexander von Humboldt foundation under the Humboldt Research Fellowship framework. The authors would further like to thank the anonymous

reviewer for their elaborate and helpful comments on the article. We greatly appreciate the time and effort the reviewer spent in providing those comments.

References

G. Stefanou. The stochastic finite element method: Past, present and future. Computer Methods in Applied Mechanics and Engineering, 198(9–12) (2009) 1031–1051.

E. Vanmarcke, Random Fields: Analysis and Synthesis, MIT Press, Cambridge, 1983.

Y. Wang, T. Zhao, K.-K. Phoon. Direct simulation of random field samples from sparsely measured geotechnical data with consideration of uncertainty in interpretation. Canadian Geotechnical Journal, 55(6) (2018), 862–880.

S. Montoya-Noguera, T. Zhao, Y. Hu, Y. Wang, K.-K. Phoon. Simulation of non-stationary non-Gaussian random fields from sparse measurements using Bayesian compressive sampling and Karhunen-Loève expansion. Structural Safety, 79 (2019), 66–79.

Y. Wang, T. Zhao, Y. Hu, K.-K. Phoon. Simulation of Random Fields with Trend from Sparse Measurements without Detrending. Journal of Engineering Mechanics, 145(2) (2019), 04018130.

J. Ching, K.-K. Phoon. Constructing a Site-Specific Multivariate Probability Distribution Using Sparse, Incomplete, and Spatially Variable (MUSIC-X) Data. Journal of Engineering Mechanics, 146(7) (2020), 1–18.

M. Beer, S. Ferson, V. Kreinovich. Imprecise probabilities in engineering analyses. Mechanical Systems and Signal Processing, 37(1–2) (2013), 4–29.

P. Spanos, R. Ghanem, Stochastic finite element expansion for random media, Journal of Engineering Mechanics 115 (5) (1989) 1035–1053.

W. Betz, I. Papaioannou, D. Straub, Numerical methods for the discretization of random fields by means of the Karhunen-Loève expansion, Computer Methods in Applied Mechanics and Engineering 271 (2014) 109–129.

M. Grigoriu, Probabilistic models for stochastic elliptic partial differential equations, Journal of Computational Physics 229 (22) (2010) 8406 – 8429.

S. P. Huang, S. T. Quek, K. K. Phoon, Convergence study of the truncated karhunen–loeve expansion for simulation of stochastic processes, International Journal for Numerical Methods in Engineering 52 (9) (2001) 1029–1043.

H. Kim, M. D. Shields, Modeling strongly non-Gaussian non-stationary stochastic processes using the Iterative Translation Approximation Method and Karhunen-Loève expansion, Computers and Structures 161 (2015) 31–42.

M.M. Dannert, A. Fau, R.M.N Fleury, M. Broggi, U. Nackenhorst, M. Beer. A probability-box approach on uncertain correlation lengths by stochastic finite element method. PAMM (Proceedings in Applied Mathematics and Mechanics), 18(1) (2018) e201800114.

M. Faes, D. Moens. Imprecise random field analysis with parametrized kernel functions. Mechanical Systems and Signal Processing, 134 (2019), 106334.

M. Fina, P. Weber, W. Wagner. Polymorphic uncertainty modeling for the simulation of geometric imperfections in probabilistic design of cylindrical shells. Structural Safety, 82 (2020), 101894.

M. Grigoriu, Simulation of stationary non-Gaussian translation processes, Journal of Engineering Mechanics 124 (2) (1998) 121–126.

M. Faes, D. Moens, On auto- and cross-interdependence in interval field finite element analysis, International Journal for Numerical Methods in Engineering 121 (9) (2020) 2033–2050.

M. Faes, D. Moens, Recent Trends in the Modeling and Quantification of Non-probabilistic Uncertainty, Archives of Computational Methods in Engineering 27 (3) (2020), 633–671.

F.N. Fritsch, R.E. Carlson. Monotone Piecewise Cubic Interpolation. SIAM Journal on Numerical Analysis 17(2) (1980), 238–246.

M.A. Misraji, M.A. Valdebenito, H.A. Jensen, C.F. Mayorga, Application of directional importance sampling for estimation of first excursion probabilities of linear structural systems subject to stochastic Gaussian loading. Mechanical Systems and Signal Processing, 139 (2020) 106621.

M.G.R. Faes, M.A. Valdebenito, D. Moens, M. Beer. Bounding the first excursion probability of linear structures subjected to imprecise stochastic loading. Computers & Structures, 239 (2020) 106320.

M.G.R. Faes, M.A. Valdebenito, X. Yuan, P. Wei, M. Beer. Augmented Reliability Analysis for Estimating Imprecise First Excursion Probabilities in Stochastic Linear Dynamics (2020). Preprint Submitted to Elsevier.

S.-K. Au, E. Patelli. Rare event simulation in finite-infinite dimensional space. Reliability Engineering & System Safety, 148 (2016) 67–77.

R. Schöbi, B. Sudret. Uncertainty propagation of p-boxes using sparse polynomial chaos expansions. Journal of Computational Physics, 339 (2017) 307–327.