Structural Safety 108 (2024) 102445

Contents lists available at ScienceDirect

Structural Safety

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

Control variates with splitting for aggregating results of Monte Carlo simulation and perturbation analysis

Cristóbal H. Acevedo [a], Marcos A. Valdebenito [b,*], Iván V. González [c], Héctor A. Jensen [d], Matthias G.R. Faes [b], Yong Liu [e]

[a] Consulting Engineer, San Antonio, Chile
[b] Chair for Reliability Engineering, TU Dortmund University, Leonhard-Euler-Strasse 5, 44227 Dortmund, Germany
[c] Department of Operational Intelligence, Esval, Valparaíso, Chile
[d] Universidad Tecnica Federico Santa Maria, Dept. de Obras Civiles, Av. España 1680, Valparaíso, Chile
[e] State Key Laboratory of Water Resources Engineering and Management, Wuhan University, Wuhan 430072, PR China

ARTICLE INFO

Keywords:
Stochastic finite element model
Second order statistics
Monte Carlo simulation
Perturbation
Intervening variables
Control variates
Splitting

ABSTRACT

Estimation of second-order statistics allows characterizing the uncertainty associated with the response of stochastic finite element models. Two common approaches for estimating these statistics are Monte Carlo simulation and perturbation. The purpose of this paper is to present a framework to aggregate the results obtained by means of these two approaches under the umbrella of Control Variates with Splitting. This allows to produce estimates of the second-order statistics of the system's response with improved precision and accuracy. More specifically, Control Variates is implemented in such a way that the variance of the estimates of second-order statistics is minimized. In addition, the application of intervening variables for enhancing perturbation is considered as well, showing substantial advantages by increasing the accuracy of the estimates of second-order statistics. The application of the proposed framework is illustrated by means of an example involving the estimation of second-order statistics of a model involving confined seepage flow.

1. Introduction

Stochastic finite element models allow representing and capturing the behavior of systems whose input parameters are uncertain [1–3]. Such task is accomplished by explicitly characterizing the uncertainties associated with system's properties or loading by means of appropriate uncertainty models [4–6]. In particular, random fields allow representing uncertainties that involve a spatial component [7–10]. Thus, stochastic finite element analysis provides a powerful tool for modeling and quantifying the behavior of engineering systems. Although analytical solutions exist for certain classes of problems (see, e.g. [11,12]), the practical deployment of stochastic finite element analysis may become challenging as in more general cases, no closed-form solutions exist. Therefore, a number of specific approaches have been developed for coping with stochastic finite element models, such as the Monte Carlo method and its advanced variants (see, e.g. [13–15]), perturbation methods (see, e.g. [16–18]), spectral methods (see, e.g. [1,19]), variability response functions (see, e.g. [12,20]), and probability density methods (see, e.g. [21,22]), to name a few.

As quantifying the uncertainty associated with a stochastic finite element model is a challenging task, a common approach is characterizing its response in terms of its second-order statistics. In general, second-order statistics may provide limited information in comparison to other probabilistic descriptors such as higher-order moments [18] and probability estimates. However, these statistics can still provide valuable insight for the problem at hand, particularly because its calculation is commonly less involved from a numerical viewpoint than tail probabilities, see e.g. [23,24]. Among the various existing approaches for calculating second-order statistics (see, e.g. [12,25–27]), two classical approaches are perturbation analysis and Monte Carlo simulation. Perturbation analysis (see, e.g. [17]) is based on constructing a linear expansion of the system's response with respect to the uncertain input parameters. Based on this linear expansion, it is possible to estimate second-order statistics of the response in closed-form. In this sense, the estimates produced by perturbation are regarded as *precise*,[1] as they are a deterministic quantity. Nonetheless, these estimates may lack *accuracy*,[1] as their value may be different from the true one due to the linearization of the response. However, this shortcoming can be overcome by resorting to intervening

* Corresponding author.
E-mail address: marcos.valdebenito@tu-dortmund.de (M.A. Valdebenito).

[1] Appendix A discusses the concepts of accuracy and precision in more detail.

https://doi.org/10.1016/j.strusafe.2024.102445
Received 30 August 2023; Received in revised form 11 January 2024; Accepted 18 January 2024
Available online 20 January 2024

variables [28], which significantly enhance the range of application of classical perturbation [29]. Still, the application of perturbation methods (with or without considering intervening variables) produces precise estimates of the sought statistics with no information on the approximation error. As a counterpart, Monte Carlo simulation (see, e.g. [13]) offers the possibility of producing *accurate* estimates of the second-order statistics, which are calculated based on random samples of the response. Nonetheless, the *precision* of these estimates depends on the number of generated samples. In fact, high precision levels can be achieved at the expense of a considerable number of random samples. The level of precision of Monte Carlo estimates can be quantified, for example, in terms of their variance by post-processing the associated random samples.

The preceding discussion highlights the advantages and drawbacks of both perturbation analysis and Monte Carlo simulation for estimating second-order statistics. A possible means for exploiting the best out of these two approaches is aggregating them by means of *Control Variates* (see, e.g. [30]). Control Variates takes advantage of the approximate results by constructing an estimator of the second-order statistics which possesses reduced variability when compared with its plain Monte Carlo counterpart. The application of Control Variates has been studied in the literature for different applications (see e.g. [31–34]) and is closely related to other approaches such as multilevel Monte Carlo (see, e.g. [31,35–39]).

This work explores the application of Control Variates as a means for calculating second-order statistics of the response of stochastic finite element models by aggregating the results produced with Monte Carlo simulation and perturbation analysis (the latter without and with intervening variables). The focus is on investigating its application for a class of problems involving linear systems whose properties are described by means of a random field. In particular, this work possesses two novel features. First, the calculation of both mean value and variance of the stochastic response is addressed. In contrast, most contributions dealing with Control Variates focus on calculating mean values only (see, e.g. [32,33]). Second, the application of a Splitting technique is included in the analysis, as proposed by [40]. This Splitting technique allows circumventing some issues related with the practical implementation of the Control Variates estimator while avoiding bias in the estimates, which is an issue frequently overlooked in the literature, see e.g. [33,41]. The main contribution of this work is providing the means to *aggregate* the results produced by two different methods for stochastic finite element analysis. In other words, it is assumed that results from perturbation analysis (with or without considering intervening variables) and Monte Carlo simulation are already available and the objective is to aggregate these results to obtain estimates of the second-order statistics with improved precision and accuracy. The aggregation of the results is performed such that the variance of the sought probabilistic descriptors (in this case, the second-order statistics) is minimized.

The rest of this contribution is organized as follows. Section 2 presents the problem under consideration as well as the basic tools for analysis, namely Monte Carlo simulation, classical perturbation and perturbation enhanced with intervening variables. Section 3 focuses on the framework of Control Variates with Splitting based on Monte Carlo and perturbation analysis. The application of the proposed approach is illustrated by means of an example in Section 4, which involves calculating the second-order statistics of a confined seepage flow problem modeled using stochastic finite elements. The paper closes with some conclusions and outlook for future work in Section 5.

2. Estimation of second-order statistics of the response of stochastic finite element models by means of Monte Carlo and perturbation

2.1. Stochastic finite element model

Consider an engineering system whose behavior is characterized as linear and which is in an steady state regime. Several practical problems fall into that category, such as linear structural mechanics, heat transfer, seepage, etc., and they can be conveniently modeled by means of the finite element method [42]. It is further assumed that a property of the system under study possesses spatial uncertainty and that this property is described in terms of a random field. This random field is represented through an appropriate method such as the Karhunen-Loève expansion (see e.g. [43]) and discretized accordingly using, e.g. the mid-point method [7]. In this way, the discrete random field becomes dependent on a random variable vector that groups a total of n_θ random variables, that is $\boldsymbol{\Theta} = [\Theta_1, \dots, \Theta_{n_\theta}]^T$, where $(\cdot)^T$ denotes transpose of the argument. Without loss of generality, it is assumed that the probability density function $p_{\boldsymbol{\Theta}}(\boldsymbol{\theta})$ associated with this random variable vector follows a multivariate standard Gaussian distribution, where $\boldsymbol{\theta}$ denotes a realization of $\boldsymbol{\Theta}$. In such a context, a non-Gaussian random field can be modeled by a memory-less translation from an underlying Gaussian random field, see e.g. [8,44–46].

Under the assumptions described above, the behavior of the system under study can be represented in a discrete way by means of the following system of linear equations:

$$\boldsymbol{K}(\boldsymbol{\theta})\,\boldsymbol{u}(\boldsymbol{\theta}) = \boldsymbol{f} \tag{1}$$

where $\boldsymbol{K}$ is a $n_d \times n_d$ matrix that is related with the system's properties; n_d denotes the number of degrees-of-freedom of the finite element model; $\boldsymbol{u}$ and $\boldsymbol{f}$ are $n_d \times 1$ vectors that represent the system's response and external action over the system, respectively. Note that in Eq. (1), it is assumed that the system's matrix $\boldsymbol{K}(\boldsymbol{\theta})$ is affected by the uncertainty associated with the random field while the external action $\boldsymbol{f}$ is deterministic. The latter is a simplification, as in reality the external actions are usually subject to uncertainty. Nonetheless, this simplified representation is retained throughout this contribution to study the effect of system's properties uncertainties on its response [17].

2.2. Second-order statistics of the response

The characterization of the system's behavior in terms of Eq. (1) allows determining a certain *response of interest* $r(\boldsymbol{\theta})$, which is assumed to be a scalar quantity and a known function of the system's response, that is $r(\boldsymbol{\theta}) = h(\boldsymbol{u}(\boldsymbol{\theta}))$. For example, the function $h(\cdot)$ may return a particular component of the response vector, an average over a number of responses at different degrees-of-freedom, etc. It is noted that the response of interest r is uncertain, as it is a function of the system's response $\boldsymbol{u}(\boldsymbol{\theta})$, which in turn depends on the particular realization $\boldsymbol{\theta}$ of the input random variable vector $\boldsymbol{\Theta}$ associated with the representation of the random field. Thus, it is of interest quantifying the uncertainty associated with this response r in terms of its mean μ_1' and variance μ_2, which are defined as [30]:

$$\mu_1'(r(\boldsymbol{\theta})) = \int_{\boldsymbol{\theta}\in\mathcal{R}^{n_\theta}} r(\boldsymbol{\theta}) p_{\boldsymbol{\Theta}}(\boldsymbol{\theta}) d\boldsymbol{\theta} \tag{2}$$

$$\mu_2(r(\boldsymbol{\theta})) = \int_{\boldsymbol{\theta}\in\mathcal{R}^{n_\theta}} (r(\boldsymbol{\theta}) - \mathbb{E}[r(\boldsymbol{\theta})])^2 p_{\boldsymbol{\Theta}}(\boldsymbol{\theta}) d\boldsymbol{\theta} \tag{3}$$

where $\mathcal{R}$ represents the set of real numbers; and $\mathbb{E}[\cdot]$ denotes expectation of the argument. Note that the notation μ_1' is consistent with that of the first order moment of response about zero (which is actually the mean) while μ_2 denotes the second order moment about the mean (which is actually the variance). In other words, the notation μ_p' denotes the pth moment about zero while μ_p denotes the pth moment about the mean [47].

For cases of practical interest, solutions for the integrals in Eqs. (2) and (3) are not available in closed form because, e.g. the dimensionality n_θ of the vector of random variables may be considerable (thus preventing the application of traditional quadrature rules) or the response $r(\boldsymbol{\theta})$ is not available in closed-form. Two possible approaches that allow circumventing these issues are Monte Carlo simulation and perturbation, which are described in Sections 2.3 and 2.4, respectively.

As an additional remark, it should be noted that in several cases, one may be interested in calculating the second-order statistics of a vector-valued system's response. In such case, the aim is estimating both the mean vector and the covariance matrix of the response [17]. Such case, which can also be treated in the context of the present formulation, is not pursued further in this work for the sake of simplicity.

2.3. Monte Carlo simulation

Monte Carlo simulation (see, e.g. [13,14,30]) offers a feasible means for estimating second-order statistics. For such purpose, n independent, identically distributed realizations of the random variable vector $\boldsymbol{\Theta}$ are generated and collected in matrix $\boldsymbol{\Theta}_n$, where the ith realization is denoted as $\boldsymbol{\theta}^{(i)}$. That is, $\boldsymbol{\Theta}_n = \left[\boldsymbol{\theta}^{(1)}, \dots, \boldsymbol{\theta}^{(n)}\right]$. Then, let μ_1' denote the mean and μ_2 denote the variance of the response of interest. Unbiased estimators for these quantities are [30]:

$$\widehat{\mu_1'}\left(r, \boldsymbol{\Theta}_n\right) = \frac{1}{n}\sum_{i=1}^{n} r\left(\boldsymbol{\theta}^{(i)}\right) \tag{4}$$

$$\widehat{\mu_2}\left(r, \boldsymbol{\Theta}_n\right) = \frac{1}{n-1}\sum_{i=1}^{n}\left(r\left(\boldsymbol{\theta}^{(i)}\right) - \frac{1}{n}\sum_{i_1=1}^{n} r\left(\boldsymbol{\theta}^{(i_1)}\right)\right)^2 \tag{5}$$

where $\widehat{(\cdot)}$ denotes an estimator. As the quantities in Eqs. (4) and (5) are estimators, they possess a variance, which is quantified by means of the following expressions [48]:

$$\widehat{\sigma^2}\left[\widehat{\mu_1'}\left(r, \boldsymbol{\Theta}_n\right)\right] = \frac{\widehat{\mu_2}\left(r, \boldsymbol{\Theta}_n\right)}{n} \tag{6}$$

$$\widehat{\sigma^2}\left[\widehat{\mu_2}\left(r, \boldsymbol{\Theta}_n\right)\right] = \frac{\widehat{\mu_4}\left(r, \boldsymbol{\Theta}_n\right)}{n} - \frac{(n-3)\,\widehat{\mu_2^2}\left(r, \boldsymbol{\Theta}_n\right)}{(n-1)\,n} \tag{7}$$

where $\widehat{\sigma^2}[\cdot]$ denotes an estimator of the variance of the argument; and $\widehat{\mu_4}\left(r, \boldsymbol{\Theta}_n\right)$ is the estimator of the fourth-order central moment. Unbiased expressions of $\widehat{\mu_4}\left(r, \boldsymbol{\Theta}_n\right)$ and $\widehat{\mu_2^2}\left(r, \boldsymbol{\Theta}_n\right)$ are given in Appendix C (see Eqs. (C.6) and (C.9) and consider $q = 0$). Clearly, the variance associated with the estimates of the mean and variance of the response as shown in Eqs. (6) and (7), respectively, tends to decrease as the number of samples n increases. Although estimators with low variability are desirable, this is often not possible in practice, as considerable numerical efforts may be required for solving the system's equation (see Eq. (1)) when the number n of realizations of the input parameters is large.

2.4. Perturbation analysis

An alternative approach to calculate second-order statistics consists of approximating the system's response r as an explicit function of $\boldsymbol{\theta}$ by means of a first-order Taylor series (see, e.g. [18,49–51]). Thus, the first-order expansion of the response (denoted as r^{L}) about $\boldsymbol{\theta}^{(0)} = [0, \dots, 0]^T$ becomes equal to:

$$r(\boldsymbol{\theta}) \approx r^{\mathrm{L}}(\boldsymbol{\theta}) = r\left(\boldsymbol{\theta}^{(0)}\right) + \sum_{k=1}^{n_\theta} r_{,k}\,\theta_k \tag{8}$$

where θ_k is the kth element of $\boldsymbol{\theta}$; $r\left(\boldsymbol{\theta}^{(0)}\right)$ denotes the response of the system evaluated at $\boldsymbol{\theta}^{(0)}$; and $r_{,k} = \partial r/\partial\theta_k\big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^{(0)}}$, $k = 1, \dots, n_\theta$ denotes the partial derivative of the response with respect to the kth random variable evaluated at $\boldsymbol{\theta}^{(0)}$. Note that the latter partial derivatives can be calculated with no additional system's analyses other than the one associated with the evaluation of $r(\boldsymbol{\theta}^{(0)})$ [52]. Furthermore, analytic expressions are available in the literature for evaluating these partial derivatives, see e.g. [52,53].

Considering the linear approximation of the response as cast in Eq. (8), the mean value and variance of the system's response can be calculated in closed-form, that is [17]:

$$\mu_1'\left(r(\boldsymbol{\theta})\right) \approx \mu_1'\left(r^{\mathrm{L}}(\boldsymbol{\theta})\right) = r\left(\boldsymbol{\theta}^{(0)}\right) \tag{9}$$

$$\mu_2\left(r(\boldsymbol{\theta})\right) \approx \mu_2\left(r^{\mathrm{L}}(\boldsymbol{\theta})\right) = \sum_{k=1}^{n_\theta} r_{,k}^2 \tag{10}$$

Note that the approximate values for the mean and variance of the response as shown in Eqs. (9) and (10) are *precise*, in the sense that they are deterministic numbers. However, they may not be *accurate*, as they are derived from an approximation of the response. In practice, it has been observed that the approximations in Eqs. (9) and (10) provide accurate estimates whenever the levels of uncertainty associated with the random field are sufficiently small.

2.5. Enhancement of perturbation analysis with intervening variables

A natural path for improving the accuracy of the second-order statistics described above is increasing the order of the Taylor expansion. Nonetheless, as discussed in [17,53], the numerical efforts associated with constructing a second-order Taylor expansion may be considerable due to the necessity of calculating the Hessian of the response, while not bringing substantial accuracy improvements for the sought statistics. A different path for producing more accurate estimates of the second-order statistics of the response consists of constructing a first-order Taylor expansion of the response with respect to *intervening variables* [54]. In essence, an intervening variable is a nonlinear function of the basic random variables of a problem. A flexible and popular intervening variable is the so-called exponential one [29,55,56], which is defined as $y_k = e^{m_k\theta_k}$, $k = 1, \dots, n_\theta$, where y_k is the intervening variable and m_k is a real constant. As shown in detail in [29], the first-order Taylor expansion of the response with respect to this exponential intervening variable (which is denoted as r^{I}) is:

$$r(\boldsymbol{\theta}) \approx r^{\mathrm{I}}(\boldsymbol{\theta}) = r\left(\boldsymbol{\theta}^{(0)}\right) + \sum_{k=1}^{n_\theta} r_{,k}\left(\frac{e^{m_k\theta_k} - 1}{m_k}\right). \tag{11}$$

The exponent m_k, $k = 1, \dots, n_\theta$ can be selected by taking into account the diagonal terms of the Hessian matrix associated with the system's response. Appendix B provides a detailed explanation about how this exponent is calculated.

Considering the approximation in Eq. (11), the second-order statistics of the system's response are given by the following closed-form expressions.

$$\mu_1'\left(r(\boldsymbol{\theta})\right) \approx \mu_1'\left(r^{\mathrm{I}}(\boldsymbol{\theta})\right) = r\left(\boldsymbol{\theta}^{(0)}\right) + \sum_{k=1}^{n_\theta} r_{,k}\left(\frac{e^{m_k^2/2} - 1}{m_k}\right) \tag{12}$$

$$\mu_2\left(r(\boldsymbol{\theta})\right) \approx \mu_2\left(r^{\mathrm{I}}(\boldsymbol{\theta})\right) = \sum_{k=1}^{n_\theta} r_{,k}^2\left(\frac{e^{2m_k^2} - e^{m_k^2}}{m_k^2}\right) \tag{13}$$

Results reported in [29] suggest that the application of exponential intervening variables for estimating second-order statistics as shown in Eqs. (12) and (13) offers considerably accuracy improvements when compared to estimates produced with classical perturbation as shown in Eqs. (9) and (10). The latter statement becomes more notorious in problems where the associated random field possesses significant uncertainty levels.

3. Control variates with splitting

3.1. General remarks

Section 2 has presented two of the most common approaches for estimating second-order statistics of stochastic finite element models, namely Monte Carlo simulation and perturbation analysis (both in its classical formulation and also considering intervening variables). The estimates produced by Monte Carlo simulation are deemed as accurate, as they are based on n samples of the system's response of interest. The level of precision of these estimates will depend upon the number of simulations n, where a higher number of simulations ensures a higher precision at the expense of repeated solution of the equilibrium equation (see Eq. (1)) for the n different samples of the random variable

vector. In this context, the level of precision can be quantified in terms of, e.g. the variance of the estimators (see Eqs. (6) and (7)). On the contrary, the results produced by perturbation analysis are *precise*, as they are deterministic quantities derived from the approximate representation of the response. However, the second-order statistics produced with perturbation analysis most likely lack accuracy, as they are deduced based on an approximation of the system's response. While accuracy can be improved resorting to intervening variables as described in Section 2.5, the results produced by a perturbation approach usually do not provide any measure on the approximation error.

The above discussion highlights the individual advantages of Monte Carlo simulation and perturbation analysis to estimate second-order statistics. A possible means to exploit these approaches synergistically consists of resorting to Control Variates (see, e.g. [30]), which is the main topic of this Section. Indeed, details about Control Variates are discussed in Section 3.2. Furthermore, for the practical deployment of Control Variates, it is required to compute a so-called control parameter. Such parameter can be conveniently calculated within a Splitting scheme [40], as shown in Section 3.3. Then, Sections 3.4 and 3.5 provide detailed expressions for calculating the mean and variance of the system's response by employing Control Variates with Splitting. Some practical implementation aspects regarding the implementation of the framework for estimating second-order statistics of the system's response of stochastic finite elements are discussed in Section 3.6.

Before presenting Control Variates, it is useful to define $\tilde{r}$, which represents an approximation of the response. In this contribution, $\tilde{r}$ can refer to either the linear approximation of the response r^{L} as shown in Eq. (8) or the approximation considering intervening variables r^{I} as shown in Eq. (11).

3.2. Control variates

Assume that μ can represent either mean value μ_1' or variance μ_2. Then, the Control Variates estimator of μ is equal to [30]:

$$\hat{\mu}^{\mathrm{CV}}\left(r,\tilde{r},\boldsymbol{\Theta}_n\right)=\hat{\mu}\left(r,\boldsymbol{\Theta}_n\right)-\gamma\hat{\mu}\left(\tilde{r},\boldsymbol{\Theta}_n\right)+\gamma\mu\left(\tilde{r}\right) \tag{14}$$

where $\hat{\mu}^{\mathrm{CV}}\left(r,\tilde{r},\boldsymbol{\Theta}_n\right)$ represents the Control Variates estimate of the statistic of the system's response r calculated using n samples $\boldsymbol{\Theta}_n$ drawn using Monte Carlo and the approximate response $\tilde{r}$; $\hat{\mu}\left(r,\boldsymbol{\Theta}_n\right)$ is a Monte Carlo estimate of the mean (see Eq. (4)) or variance (see Eq. (5)) considering the *exact* response r; $\hat{\mu}\left(\tilde{r},\boldsymbol{\Theta}_n\right)$ is analogous to the latter estimator except that it is calculated using the *approximate* response $\tilde{r}$; $\mu(\tilde{r})$ corresponds to the closed-form expression for the mean (see eq. (9) or (12)) or variance (see eq. (10) or (13)) calculated using the approximate response $\tilde{r}$; and γ is a so-called control parameter, which is a real number. In order to understand the rationale behind Eq. (14), consider for a moment that the control parameter is selected such that $\gamma=1$ (such restriction is removed later on). Under such an assumption, the Control Variates estimate can be understood as follows.

- The term $\gamma\mu(\tilde{r})$ in Eq. (14) is a *precise* estimate of the sought statistic, as it is computed in closed-form. However, it may lack accuracy, as it is calculated using an approximation of the true response.
- The term $\hat{\mu}\left(r,\boldsymbol{\Theta}_n\right)-\gamma\hat{\mu}\left(\tilde{r},\boldsymbol{\Theta}_n\right)$ can be interpreted as a *correction* over the previous value, as it computes the difference between the estimate of the statistic calculated using the exact and approximate responses, respectively.

The previous discussion highlights the fact that Control Variates as cast in Eq. (14) successfully aggregates the estimators for a statistic (either mean or variance) produced with Monte Carlo simulation (which is represented by the term $\hat{\mu}\left(r,\boldsymbol{\Theta}_n\right)$) with those produced by perturbation (which is represented by the term $\mu\left(\tilde{r}\right)$).

Fig. 1 provides a schematic representation of the Control Variates estimator in Eq. (14). This representation considers the probability density functions associated to the different estimators involved. The green arrow represents the probability density function associated with $\gamma\mu(\tilde{r})$ and corresponds to a Dirac delta, as this is a precise value. Nonetheless, such estimator is not accurate, as it is different from the sought statistic $\mu(r)$. The red curve illustrates the probability density function associated with the estimator $\hat{\mu}(r,\boldsymbol{\Theta}_n)$, which corresponds to the plain Monte Carlo estimate of the sought statistic considering the exact response. This estimator is accurate, however, it may lack sufficient precision because the sample set $\boldsymbol{\Theta}_n$ is relatively small. The purple curve represents the probability distribution associated with the estimator of the correction term $\hat{\mu}\left(r,\boldsymbol{\Theta}_n\right)-\gamma\hat{\mu}\left(\tilde{r},\boldsymbol{\Theta}_n\right)$. As it is expected that the approximate and exact responses exhibit a considerable correlation, the variance of this probability density is relatively low. The blue curve represents the probability density associated with the estimator considering Control Variates, which corresponds to the superposition of the purple and green curves. This probability density is both accurate (because it leads to the sought statistic) and precise (as it has reduced variability when compared to its plain Monte Carlo counterpart).

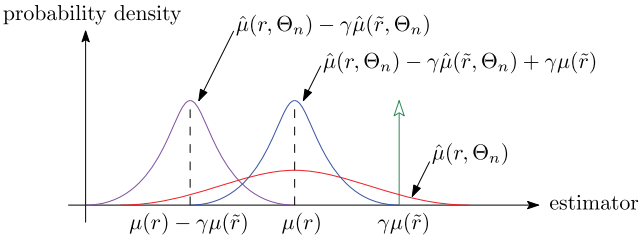

Fig. 1. Schematic representation of Control Variates.

The advantages of the Control Variates estimate as presented in Eq. (14) and Fig. 1 can be also understood by examining its variance estimate, which is equal to [30]:

$$\hat{\sigma}^2\left[\hat{\mu}^{\mathrm{CV}}\left(r,\tilde{r},\boldsymbol{\Theta}_n\right)\right]=\hat{\sigma}^2\left[\hat{\mu}\left(r,\boldsymbol{\Theta}_n\right)\right]-2\gamma\hat{\delta}\left[\hat{\mu}\left(r,\boldsymbol{\Theta}_n\right),\hat{\mu}\left(\tilde{r},\boldsymbol{\Theta}_n\right)\right]+\gamma^2\hat{\sigma}^2\left[\hat{\mu}\left(\tilde{r},\boldsymbol{\Theta}_n\right)\right] \tag{15}$$

where $\hat{\delta}[\cdot,\cdot]$ denotes the estimator of the covariance between the arguments. A close examination of Eq. (15) indicates that its first and third terms (that is, $\hat{\sigma}^2\left[\hat{\mu}\left(r,\boldsymbol{\Theta}_n\right)\right]$ and $\gamma^2\hat{\sigma}^2\left[\hat{\mu}\left(\tilde{r},\boldsymbol{\Theta}_n\right)\right]$, respectively) measure the variance associated with the first and second terms of Eq. (14) (that is, $\hat{\mu}\left(r,\boldsymbol{\Theta}_n\right)$ and $\gamma\hat{\mu}\left(\tilde{r},\boldsymbol{\Theta}_n\right)$, respectively) while the second term of Eq. (15) (that is, $2\gamma\hat{\delta}\left[\hat{\mu}\left(r,\boldsymbol{\Theta}_n\right),\hat{\mu}\left(\tilde{r},\boldsymbol{\Theta}_n\right)\right]$) appears due to the covariance between the first and second terms of Eq. (14). In addition, note that the third term of Eq. (14) (that is, $\gamma\mu(\tilde{r})$) is a deterministic quantity and hence, it does not contribute to the variance expression in Eq. (15).

It is noted that Eq. (15) is a quadratic function with respect to the control parameter γ. Thus, by removing the previous assumption that $\gamma=1$, one can select γ such that the variance of the estimator as shown in Eq. (15) is minimized. This implies taking the derivative of Eq. (15) with respect to γ and equating it to zero, which leads to the optimal value of the control parameter [30,57], which is denoted as γ^*:

$$\gamma^*=\frac{\hat{\delta}\left[\hat{\mu}\left(r,\boldsymbol{\Theta}_n\right),\hat{\mu}\left(\tilde{r},\boldsymbol{\Theta}_n\right)\right]}{\hat{\sigma}^2\left[\hat{\mu}\left(\tilde{r},\boldsymbol{\Theta}_n\right)\right]}. \tag{16}$$

As noted from the above equation, the optimal control parameter involves estimating variance and covariance of the system's response r and its approximation $\tilde{r}$. If the optimal control parameter in Eq. (16) is replaced into Eq. (15), one obtains:

$$\hat{\sigma}^2_{\min}\left[\hat{\mu}^{\mathrm{CV}}\left(r,\tilde{r},\boldsymbol{\Theta}_n\right)\right]=\hat{\sigma}^2\left[\hat{\mu}\left(r,\boldsymbol{\Theta}_n\right)\right]-\frac{\left(\hat{\delta}\left[\hat{\mu}\left(r,\boldsymbol{\Theta}_n\right),\hat{\mu}\left(\tilde{r},\boldsymbol{\Theta}_n\right)\right]\right)^2}{\hat{\sigma}^2\left[\hat{\mu}\left(\tilde{r},\boldsymbol{\Theta}_n\right)\right]} \tag{17}$$

which expresses the minimum possible variance that the Control Variates estimate may attain and which is denoted as $\hat{\sigma}^2_{\min}\left[\hat{\mu}^{\mathrm{CV}}\left(r,\tilde{r},\boldsymbol{\Theta}_n\right)\right]$.

By examining Eq. (17), it is noted that the first term represents the variance of the Monte Carlo estimator. The second term is associated with the covariance between the statistics calculated with the exact system's response r and its approximation $\tilde{r}$. Hence, provided that $\tilde{r}$ closely mimics r, it is expected that the covariance between estimates is high and therefore, the variance of the Control Variates estimate can be significantly smaller than that of its Monte Carlo counterpart. In fact, if the approximate response $\tilde{r}$ were to reproduce the response r exactly, the variance of the Control Variates estimate would be zero. Of course, the latter case just illustrates an extreme situation as in practical applications, one can expect discrepancies between $\tilde{r}$ and r.

As a summary of the above discussion, it is observed that Control Variates offers an excellent means to aggregate the results of perturbation analysis and Monte Carlo simulation for problems involving estimation of second-order statistics of stochastic finite element models. In fact, as perturbation is based on constructing an approximate representation of the response, it is expected that there is considerable covariance between the statistics estimated with the two approaches (that is, Monte Carlo and perturbation), leading to an estimate with improved accuracy (as given by Eq. (14)) and with high precision (that is, with minimum variance, as expressed in Eq. (17)). In the worst-case scenario where there is no covariance between the exact response r and the approximate response $\tilde{r}$, Control Variates would still allow to retrieve the results from classical Monte Carlo simulation. This last statement can be easily verified by examining Eq. (16): in case of null covariance, the optimal control parameter becomes $\gamma^* = 0$ and thus, Eqs. (14) and (17) become the classical Monte Carlo estimates. Furthermore, aggregating the results of perturbation and Monte Carlo simulation under the umbrella provided by Control Variates allows quantifying the approximation error associated with the estimated second-order statistics in terms of their variance. This is a significant advantage, as a perturbation approach (with or without considering intervening variables) does not usually offer an explicit error estimate.

3.3. Splitting

The previous Section has highlighted the advantages of Control Variates. However, for its practical implementation, it is necessary to estimate the optimal control parameter γ^* by means of Eq. (16), which in turn demands knowledge on the covariance and variance of the involved estimators. If this control parameter were to be calculated using the very same samples used to estimate the sought statistic as shown in Eq. (14), this would introduce bias in the estimates [33,58]. The latter issue could be particularly harmful in case that the number of samples n is small [32], which is expected to be the case in practical applications. Fortunately, the effect of bias can be conveniently avoided by the so-called *Splitting* technique, as introduced in [40]. Splitting consists of subdividing the pool of n samples of the uncertain parameters $\boldsymbol{\Theta}_n$ into k subsets. Each of these k subsets contains a total of $n^* = n/k$ samples, where it is assumed that n and k are selected such that n^* is a positive integer number. The jth sample subset is denoted as $\boldsymbol{\Theta}_{n^*,j}$ and it contains the following samples:

$$\boldsymbol{\Theta}_{n^*,j} = \left[\theta^{(1+(j-1)n^*)}, \ldots, \theta^{(n^*+(j-1)n^*)}\right], \; j = 1, \ldots, k. \tag{18}$$

In addition, consider the following integer τ as a function of the sample subset identifier j:

$$\tau(j) = \text{rem}(j,k) + 1, \; j = 1, \ldots, k \tag{19}$$

where $\text{rem}(j,k)$ returns the remainder of the division between j and k. Note that in case $k \geq 2$, it is always ensured that $\tau(j) \neq j, \; j = 1, \ldots, k$.

Taking into account the definitions in Eqs. (18) and (19), the Control Variates estimator that includes the Splitting scheme is:

$$\hat{\mu}^{\text{CV+S}}\left(r, \tilde{r}, \boldsymbol{\Theta}_n\right) = \frac{1}{k}\sum_{j=1}^{k} \hat{\mu}^{\text{CV},(j)}\left(r, \tilde{r}, \boldsymbol{\Theta}_{n^*,j}, \boldsymbol{\Theta}_{n^*,\tau(j)}\right). \tag{20}$$

In the above equation, $\hat{\mu}^{\text{CV+S}}\left(r, \tilde{r}, \boldsymbol{\Theta}_n\right)$ denotes the Control Variates estimator considering Splitting; $\hat{\mu}^{\text{CV},(j)}\left(r, \tilde{r}, \boldsymbol{\Theta}_{n^*,j}\boldsymbol{\Theta}_{n^*,\tau(j)}\right)$ is the Control Variates estimate associated with the jth and $\tau(j)$-th sample subsets, that is defined as [40]:

$$\begin{aligned}\hat{\mu}^{\text{CV},(j)}\left(r, \tilde{r}, \boldsymbol{\Theta}_{n^*,j}, \boldsymbol{\Theta}_{n^*,\tau(j)}\right) = \hat{\mu}\left(r, \boldsymbol{\Theta}_{n^*,j}\right) - \gamma^*_{\tau(j)}\left(r, \tilde{r}, \boldsymbol{\Theta}_{n^*,\tau(j)}\right)\hat{\mu}\left(\tilde{r}, \boldsymbol{\Theta}_{n^*,j}\right) \\ + \gamma^*_{\tau(j)}\left(r, \tilde{r}, \boldsymbol{\Theta}_{n^*,\tau(j)}\right)\mu\left(\tilde{r}\right), \; j = 1, \ldots, k\end{aligned} \tag{21}$$

where the optimal control parameter $\gamma^*_{\tau(j)}$ is calculated as:

$$\gamma^*_{\tau(j)}\left(r, \tilde{r}, \boldsymbol{\Theta}_{n^*,\tau(j)}\right) = \frac{\hat{\delta}\left[\hat{\mu}\left(r, \boldsymbol{\Theta}_{n^*,\tau(j)}\right), \hat{\mu}\left(\tilde{r}, \boldsymbol{\Theta}_{n^*,\tau(j)}\right)\right]}{\hat{\sigma}^2\left[\hat{\mu}\left(\tilde{r}, \boldsymbol{\Theta}_{n^*,\tau(j)}\right)\right]}, \; j = 1, \ldots, k. \tag{22}$$

As seen in Eq. (20), the estimator based on Control Variates with Splitting is simply the average of the basic Control Variates estimates $\hat{\mu}^{\text{CV},(j)}\left(r, \tilde{r}, \boldsymbol{\Theta}_{n^*,j}, \boldsymbol{\Theta}_{n^*,\tau(j)}\right), \; j = 1, \ldots, k$, shown in Eq. (21). Moreover, by examining Eqs. (21) and (22), it is observed that this basic Control Variates estimate $\hat{\mu}^{\text{CV},(j)}\left(r, \tilde{r}, \boldsymbol{\Theta}_{n^*,j}, \boldsymbol{\Theta}_{n^*,\tau(j)}\right)$ is calculated such that:

- The estimators of the statistics of the exact response r and its approximation $\tilde{r}$ are calculated with the jth sample subset, as shown in Eq. (21).
- The optimal control parameter γ^* is calculated using the $\tau(j)$-th sample subset.

This implies that, in practice, the statistics estimated with one sample subset are *controlled* by another sample subset, where *controlled* means the application of the optimal control parameter γ^*. Such strategy allows effectively avoiding the effect of bias in the Control Variates estimate as long as the number of sample subsets is chosen such that $k \geq 3$ [40]. However, selecting a number k of sample subsets which is too large increases the variance of the estimator. Therefore, for practical applications, it is recommended that $k = 3$.

The estimator of the variance associated with the estimator applying Control Variates with Splitting is [40]:

$$\hat{\sigma}^2\left[\hat{\mu}^{\text{CV+S}}\left(r, \tilde{r}, \boldsymbol{\Theta}_n\right)\right] = \frac{1}{k^2}\sum_{j=1}^{k} \hat{\sigma}^2\left[\hat{\mu}^{\text{CV},(j)}\left(r, \tilde{r}, \boldsymbol{\Theta}_{n^*,j}, \boldsymbol{\Theta}_{n^*,\tau(j)}\right)\right], \tag{23}$$

where:

$$\begin{aligned}\hat{\sigma}^2\left[\hat{\mu}^{\text{CV},(j)}\left(r, \tilde{r}, \boldsymbol{\Theta}_{n^*,j}, \boldsymbol{\Theta}_{n^*,\tau(j)}\right)\right] = \hat{\sigma}^2\left[\hat{\mu}\left(r, \boldsymbol{\Theta}_{n^*,j}\right)\right] - 2\gamma^*_{\tau(j)}\left(r, \tilde{r}, \boldsymbol{\Theta}_{n^*,\tau(j)}\right) \\ \times \hat{\delta}\left[\hat{\mu}\left(r, \boldsymbol{\Theta}_{n^*,j}\right), \hat{\mu}\left(\tilde{r}, \boldsymbol{\Theta}_{n^*,j}\right)\right] \\ + \left(\gamma^*_{\tau(j)}\left(r, \tilde{r}, \boldsymbol{\Theta}_{n^*,\tau(j)}\right)\right)^2 \\ \times \hat{\sigma}^2\left[\hat{\mu}\left(\tilde{r}, \boldsymbol{\Theta}_{n^*,j}\right)\right], \; j = 1, \ldots, k.\end{aligned} \tag{24}$$

As a summary of this Section, it is noted that the Splitting technique simply consists of performing calculations of the different estimators and control parameters over sample subsets. The advantage of such an approach is that it can effectively avoid the effect of bias. Furthermore, when comparing the estimators based on Control Variates (as shown in Section 3.2) with those that involve Splitting (as shown in the current Section), it is noted that Splitting does not demand any additional evaluations of the system's response (that is, no additional simulations of the stochastic finite element model are required). In other words, the application of Splitting does not involve any substantial numerical burden, except for the additional calculation required for keeping track of different sample subsets. Such calculation possesses a small numerical cost when compared to the efforts spent in the evaluation of the exact system's response.

3.4. Estimation of the mean of the response

Sections 3.2 and 3.3 provide a general framework for applying Control Variates with Splitting. This Section presents specific expressions for estimating the mean value of the response as well as the variance of this estimate (that is, variance of the mean). Before presenting these

expressions, it is useful to define $\mu_{p,q} = \mathbb{E}\left[(r - \mathbb{E}[r])^p(\tilde{r} - \mathbb{E}[\tilde{r}])^q\right]$, which represents the bivariate central co-moment of order (p, q) between r and $\tilde{r}$, where p and q are natural numbers (including zero) while $\mathbb{E}[\cdot]$ denotes expectation of the argument (as already defined in Section 2.2). Appendix C provides detailed expressions for estimating these bivariate central co-moments based on samples. Furthermore, please recall that $\tilde{r}$ represents an approximation of the response, which can refer to either the linear approximation of the response r^{L} or the approximation considering intervening variables r^{I}, as shown in Eqs. (8) and (11), respectively.

The estimator for the mean of the system's response considering Control Variates with Splitting is denoted as $\widehat{\mu_1'}^{\mathrm{CV+S}}$ and is given by the following expression:

$$\widehat{\mu_1'}^{\mathrm{CV+S}}\left(r, \tilde{r}, \boldsymbol{\Theta}_n\right) = \frac{1}{k}\sum_{j=1}^{k} \widehat{\mu_1'}^{\mathrm{CV},(j)}\left(r, \tilde{r}, \boldsymbol{\Theta}_{n^*,j}, \boldsymbol{\Theta}_{n^*,\tau(j)}\right). \tag{25}$$

In the above equation, $\widehat{\mu_1'}^{\mathrm{CV},(j)}$ is defined as:

$$\begin{aligned}\widehat{\mu_1'}^{\mathrm{CV},(j)}\left(r, \tilde{r}, \boldsymbol{\Theta}_{n^*,j}, \boldsymbol{\Theta}_{n^*,\tau(j)}\right) = \widehat{\mu_1'}\left(r, \boldsymbol{\Theta}_{n^*,j}\right) - \alpha^*_{\tau(j)}\left(r, \tilde{r}, \boldsymbol{\Theta}_{n^*,\tau(j)}\right)\widehat{\mu_1'}\left(\tilde{r}, \boldsymbol{\Theta}_{n^*,j}\right) \\ + \alpha^*_{\tau(j)}\left(r, \tilde{r}, \boldsymbol{\Theta}_{n^*,\tau(j)}\right)\mu_1'\left(\tilde{r}\right),\ j = 1, \ldots, k,\end{aligned} \tag{26}$$

where $\widehat{\mu_1'}\left(r, \boldsymbol{\Theta}_{n^*,j}\right)$ and $\widehat{\mu_1'}\left(\tilde{r}, \boldsymbol{\Theta}_{n^*,j}\right)$ are Monte Carlo estimates of the mean of the exact and approximate responses, respectively, which are calculated using Eq. (4); $\mu_1'(\tilde{r})$ is the mean value of the approximate response calculated in closed-form considering either the classical perturbation solution (see Eq. (9)) or intervening variables (see Eq. (12)); and $\alpha^*_{\tau(j)}$ is the optimal control parameter for estimating the mean, which is equal to:

$$\alpha^*_{\tau(j)}\left(r, \tilde{r}, \boldsymbol{\Theta}_{n^*,\tau(j)}\right) = \frac{\widehat{\mu_{1,1}}\left(r, \tilde{r}, \boldsymbol{\Theta}_{n^*,\tau(j)}\right)}{\widehat{\mu_{0,2}}\left(r, \tilde{r}, \boldsymbol{\Theta}_{n^*,\tau(j)}\right)},\ j = 1, \ldots, k, \tag{27}$$

where $\widehat{\mu_{1,1}}$ and $\widehat{\mu_{0,2}}$ denote estimators for bivariate central moments which are calculated according to Eqs. (C.3) and (C.4), respectively. The estimator of the variance of the estimate for the mean in Eq. (25) is given by:

$$\hat{\sigma}^2\left[\widehat{\mu_1'}^{\mathrm{CV+S}}\left(r, \tilde{r}, \boldsymbol{\Theta}_n\right)\right] = \frac{1}{k^2}\sum_{j=1}^{k}\hat{\sigma}^2\left[\widehat{\mu_1'}^{\mathrm{CV},(j)}\left(r, \tilde{r}, \boldsymbol{\Theta}_{n^*,j}, \boldsymbol{\Theta}_{n^*,\tau(j)}\right)\right], \tag{28}$$

where $\hat{\sigma}^2\left[\widehat{\mu_1'}^{\mathrm{CV},(j)}\left(r, \tilde{r}, \boldsymbol{\Theta}_{n^*,j}, \boldsymbol{\Theta}_{n^*,\tau(j)}\right)\right]$ is equal to:

$$\begin{aligned}\hat{\sigma}^2\left[\widehat{\mu_1'}^{\mathrm{CV},(j)}\left(r, \tilde{r}, \boldsymbol{\Theta}_{n^*,j}, \boldsymbol{\Theta}_{n^*,\tau(j)}\right)\right] = \frac{\widehat{\mu_{2,0}}\left(r, \tilde{r}, \boldsymbol{\Theta}_{n^*,j}\right)}{n^*} - 2\alpha^*_{\tau(j)}\left(r, \tilde{r}, \boldsymbol{\Theta}_{n^*,\tau(j)}\right) \\ \times \frac{\widehat{\mu_{1,1}}\left(r, \tilde{r}, \boldsymbol{\Theta}_{n^*,j}\right)}{n^*} \\ + \left(\alpha^*_{\tau(j)}\left(r, \tilde{r}, \boldsymbol{\Theta}_{n^*,\tau(j)}\right)\right)^2\frac{\widehat{\mu_{0,2}}\left(r, \tilde{r}, \boldsymbol{\Theta}_{n^*,j}\right)}{n^*}, \\ j = 1, \ldots, k,\end{aligned} \tag{29}$$

where $\widehat{\mu_{2,0}}$ is calculated according to Eq. (C.2).

3.5. Estimation of the variance of the response

This Section presents expressions for estimating the variance of the response as well as the variance of this estimate (that is, variance of the variance) based on the material presented in Sections 3.2 and 3.3. The estimator $\widehat{\mu_2}^{\mathrm{CV+S}}$ denotes the variance of the system's response considering Control Variates with Splitting and is equal to:

$$\widehat{\mu_2}^{\mathrm{CV+S}}\left(r, \tilde{r}, \boldsymbol{\Theta}_n\right) = \frac{1}{k}\sum_{j=1}^{k}\widehat{\mu_2}^{\mathrm{CV},(j)}\left(r, \tilde{r}, \boldsymbol{\Theta}_{n^*,j}, \boldsymbol{\Theta}_{n^*,\tau(j)}\right). \tag{30}$$

In the last equation, $\widehat{\mu_2}^{\mathrm{CV},(j)}$ is equal to:

$$\begin{aligned}\widehat{\mu_2}^{\mathrm{CV},(j)}\left(r, \tilde{r}, \boldsymbol{\Theta}_{n^*,j}, \boldsymbol{\Theta}_{n^*,\tau(j)}\right) = \widehat{\mu_2}\left(r, \boldsymbol{\Theta}_{n^*,j}\right) - \beta^*_{\tau(j)}\left(r, \tilde{r}, \boldsymbol{\Theta}_{n^*,\tau(j)}\right)\widehat{\mu_2}\left(\tilde{r}, \boldsymbol{\Theta}_{n^*,j}\right) \\ + \beta^*_{\tau(j)}\left(r, \tilde{r}, \boldsymbol{\Theta}_{n^*,\tau(j)}\right)\mu_2\left(\tilde{r}\right),\ j = 1, \ldots, k,\end{aligned} \tag{31}$$

where $\widehat{\mu_2}\left(r, \boldsymbol{\Theta}_{n^*,j}\right)$ and $\widehat{\mu_2}\left(\tilde{r}, \boldsymbol{\Theta}_{n^*,j}\right)$ are Monte Carlo estimates of the variance of the exact and approximate responses, respectively, which are calculated using Eq. (5); $\mu_2(\tilde{r})$ is the variance of the approximate response calculated in closed form by means of either a linear approximation (see Eq. (10)) or intervening variables (see Eq. (13)); and $\beta^*_{\tau(j)}$ is the optimal control parameter for estimating the variance. The expression for calculating this optimal control parameters is presented afterwards.

The estimator of the variance of the estimate for the variance in Eq. (30) is given by:

$$\hat{\sigma}^2\left[\widehat{\mu_2}^{\mathrm{CV+S}}\left(r, \tilde{r}, \boldsymbol{\Theta}_n\right)\right] = \frac{1}{k^2}\sum_{j=1}^{k}\hat{\sigma}^2\left[\widehat{\mu_2}^{\mathrm{CV},(j)}\left(r, \tilde{r}, \boldsymbol{\Theta}_{n^*,j}, \boldsymbol{\Theta}_{n^*,\tau(j)}\right)\right], \tag{32}$$

where $\hat{\sigma}^2\left[\widehat{\mu_2}^{\mathrm{CV},(j)}\left(r, \tilde{r}, \boldsymbol{\Theta}_{n^*,j}, \boldsymbol{\Theta}_{n^*,\tau(j)}\right)\right]$ is defined as:

$$\begin{aligned}\hat{\sigma}^2\left[\widehat{\mu_2}^{\mathrm{CV},(j)}\left(r, \tilde{r}, \boldsymbol{\Theta}_{n^*,j}, \boldsymbol{\Theta}_{n^*,\tau(j)}\right)\right] = B_1\left(r, \tilde{r}, \boldsymbol{\Theta}_{n^*,j}\right) - 2\beta^*_{\tau(j)}\left(r, \tilde{r}, \boldsymbol{\Theta}_{n^*,\tau(j)}\right) \\ \times B_2\left(r, \tilde{r}, \boldsymbol{\Theta}_{n^*,j}\right) \\ + \left(\beta^*_{\tau(j)}\left(r, \tilde{r}, \boldsymbol{\Theta}_{n^*,\tau(j)}\right)\right)^2 \\ \times B_3\left(r, \tilde{r}, \boldsymbol{\Theta}_{n^*,j}\right),\ j = 1, \ldots, k,\end{aligned} \tag{33}$$

where $B_1\left(r, \tilde{r}, \boldsymbol{\Theta}_{n^*,j}\right)$, $B_2\left(r, \tilde{r}, \boldsymbol{\Theta}_{n^*,j}\right)$ and $B_3\left(r, \tilde{r}, \boldsymbol{\Theta}_{n^*,j}\right)$ are defined as:

$$B_1\left(r, \tilde{r}, \boldsymbol{\Theta}_{n^*,j}\right) = \frac{\widehat{\mu_{4,0}}\left(r, \tilde{r}, \boldsymbol{\Theta}_{n^*,j}\right)}{n^*} - \frac{(n^*-3)}{(n^*-1)n^*}\widehat{\mu^2_{2,0}}\left(r, \tilde{r}, \boldsymbol{\Theta}_{n^*,j}\right) \tag{34}$$

$$B_2\left(r, \tilde{r}, \boldsymbol{\Theta}_{n^*,j}\right) = \frac{2\widehat{\mu^2_{1,1}}\left(r, \tilde{r}, \boldsymbol{\Theta}_{n^*,j}\right)}{(n^*-1)n^*} + \frac{\widehat{\mu_{2,2}}\left(r, \tilde{r}, \boldsymbol{\Theta}_{n^*,j}\right)}{n^*} - \frac{\widehat{\mu_{2,0}\mu_{0,2}}\left(r, \tilde{r}, \boldsymbol{\Theta}_{n^*,j}\right)}{n^*} \tag{35}$$

$$B_3\left(r, \tilde{r}, \boldsymbol{\Theta}_{n^*,j}\right) = \frac{\widehat{\mu_{0,4}}\left(r, \tilde{r}, \boldsymbol{\Theta}_{n^*,j}\right)}{n^*} - \frac{(n^*-3)}{(n^*-1)n^*}\widehat{\mu^2_{0,2}}\left(r, \tilde{r}, \boldsymbol{\Theta}_{n^*,j}\right) \tag{36}$$

and estimators for all the bivariate central co-moments that appear in the above equations can be found in Appendix C. Lastly, the estimator for the optimal control parameter $\beta^*_{\tau(j)}$ is given by the following expression.

$$\beta^*_{\tau(j)}\left(r, \tilde{r}, \boldsymbol{\Theta}_{n^*,\tau(j)}\right) = \frac{B_2\left(r, \tilde{r}, \boldsymbol{\Theta}_{n^*,\tau(j)}\right)}{B_3\left(r, \tilde{r}, \boldsymbol{\Theta}_{n^*,\tau(j)}\right)},\ j = 1, \ldots, k. \tag{37}$$

Note that the terms B_2 and B_3 in Eq. (37) are calculated using the expressions in eqs. (35) and (36), respectively. However, in this case, one should consider the set of samples $\boldsymbol{\Theta}_{n^*,\tau(j)}$ instead of $\boldsymbol{\Theta}_{n^*,j}$, in accordance with the Splitting technique.

3.6. Practical implementation

The practical implementation of Control Variates with Splitting as described in this Section is carried out under the assumption that results from perturbation analysis (with or without considering intervening variables) and Monte Carlo simulation are already available, as already stated in Section 1. Thus, the objective is to aggregate the results already produced by each of these two methods. The main steps involved in the proposed approach under the assumption described previously are the following.

1. Approximate the response of the stochastic finite element model and calculate the corresponding second-order statistics. Use Eqs. (8), (9) and (10) in case that the response is approximated using a first-order Taylor expansion; or Eqs. (11), (12) and (13) in case that intervening variables are considered.
2. Conduct Monte Carlo simulation. Consider a total of n samples. Calculate for each sample both the exact response $r\left(\boldsymbol{\theta}^{(i)}\right)$, $i = 1, \ldots, n$ and the approximate response $\tilde{r}\left(\boldsymbol{\theta}^{(i)}\right)$, $i = 1, \ldots, n$. Note that the approximate response must be calculated with the

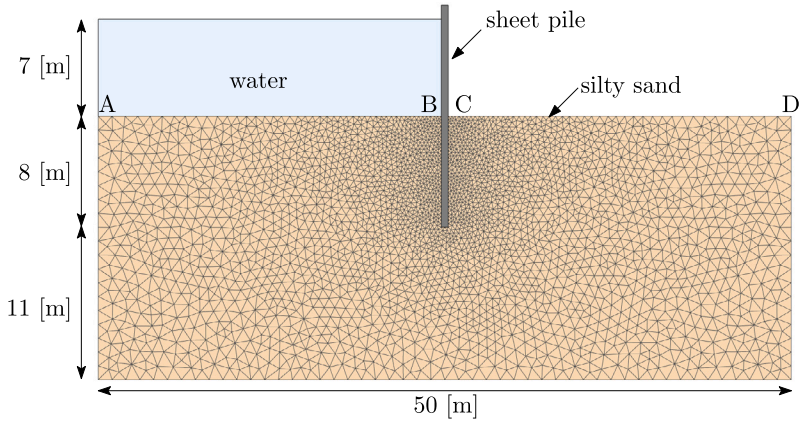

Fig. 2. Example – Seepage below sheet pile (schematic representation).

approximation selected in the previous step (either first-order Taylor expansion or intervening variables).

3. Split the sample set generated in the previous step into $k = 3$ subsets following Eq. (18).
4. Apply Eqs. (25) and (28) to estimate the mean of the response and the variance of mean using Control Variates with splitting.
5. Apply Eqs. (30) and (32) to estimate the variance of the response and the variance of variance using Control Variates with splitting.

For the practical implementation of the framework described above, it is recommended that the number of samples n associated with Monte Carlo is such that $n \geq 60$. Such criterion ensures that, at least, there are 20 samples in each of the subsets required for the splitting step. Numerical validations suggest that such criterion ensures appropriate estimation of the sought statistics as well as the required co-moments.

4. Example

4.1. Description

The application of the framework developed in Section 3 is illustrated here by means of the following example, which consists of quantifying the second-order statistics associated with the seepage flow beneath the sheet pile illustrated schematically in Fig. 2. This class of problems is quite challenging from the point of view of uncertainty quantification, as documented in e.g., [59].

The sheet pile retains a water column of 7 [m] height and rests over a silty sand soil layer. The permeability of this soil layer is modeled as a log-normal random field, with mean value 5×10^{-6} [m/s] and standard deviation 5×10^{-6} [m/s]. In this context, note that the assumption of a log-normal random field is consistent with the physics of the problem at hand, that is, that the permeability admits positive values only (see, e.g. [59]). The correlation is modeled using the quadratic exponential function such that e^{-d^2/L^2}, where d expresses the Euclidean distance between two points of the random field and L is the correlation length, which is taken as $L = 10$ [m]. Note that the quadratic exponential model for correlation is considered in here for simplicity, although alternative models have been promoted in the literature, see e.g. [60,61]. It is observed that the random field associated with the permeability possesses a high degree of uncertainty, as its coefficient of variation is equal to 100%. This is consistent with the high variability observed for this property in practical applications, see e.g. [62]. Furthermore, it should be noted that in practice, the vertical and horizontal permeability of a soil are usually distinct, see e.g. [63]. However, in this contribution, the permeability is modeled as isotropic for the sake of simplicity.

The seepage flow is quantified by solving the associated Laplace equation, which is discretized by means of the finite element method, see e.g., [42]. For this purpose, a mesh comprising 6090 quadratic triangular elements and 12427 nodes is considered, as illustrated schematically in Fig. 2. The boundary conditions are prescribed water heads of 7 [m] and 0 [m] over lines AB and CD in Fig. 2 (that is, a Dirichlet boundary condition), respectively, while a condition of null flow is imposed over all other boundaries (that is, a Neumann boundary condition). The log-normal random field associated with the permeability is discretized considering the mid-point method [7] and it is represented using the Karhunen-Loève expansion, where 95% of the total variability of the underlying Gaussian field is retained. This leads to a discrete representation involving a total of $n_\theta = 10$ standard normal random variables. The seepage flow and its statistics are approximated considering both the linear expansion in Eq. (8) as well as intervening variables in Eq. (11).

4.2. Numerical results

The second-order statistics of the seepage flow are estimated herein and all results are summarized in Table 1. Note that the physical units associated with the different statistics consider that seepage flow is measured as flow ([m^3/s]) over unit length of the sheet pile ([m]). The first column of this table indicates the approach considered while the second reports the total number of simulations n required for implementing a specific approach. The third column contains the total numerical cost required for implementing a certain approach in terms of the number of equivalent simulations n_E. This number summarizes the total number of simulations required in a certain approach plus costs associated with tasks such as e.g., calculation of first- or second-order sensitivities, application of Splitting technique, etc. The equivalent number of simulations n_E is calculated as:

$$n_E = n + \frac{t_A}{t_S} \tag{38}$$

where t_A denotes the time spent in additional tasks and t_S is the time spent in one simulation, that is, the assembly and solution of the equilibrium equation as reported in Eq. (1). The rest of the columns of Table 1 report the estimated statistics (either $\widehat{\mu_1'}$ or $\widehat{\mu_2}$, see fourth and seventh columns, respectively), the variance of these statistics (either $\hat{\sigma}^2\left[\widehat{\mu_1'}\right]$ or $\hat{\sigma}^2\left[\widehat{\mu_2}\right]$, see fifth and eighth columns, respectively) and the associated coefficient of variation (either $\mathrm{CoV}\left[\widehat{\mu_1'}\right]$ or $\mathrm{CoV}\left[\widehat{\mu_2}\right]$, see sixth and ninth columns, respectively). Note that the latter is equal to the square root of the variance of the estimator divided by the estimator.

The second row of Table 1 reports the second-order statistics of seepage flow estimated using Monte Carlo simulation (MCS) considering a total of $n = 3000$ samples. This is a relatively large number of

Table 1
Estimation of second-order statistics of seepage flow. MCS: Monte Carlo simulation. Lin: perturbation analysis considering linear approximation. IV: perturbation analysis considering intervening variables. CV+S: Control Variates with Splitting.

Approach	n	n_E	$\widehat{\mu_1'}$ [m³/s/m]	$\widehat{\sigma}^2\left[\widehat{\mu_1'}\right]$ [m⁶/s²/m²]	CoV$\left[\widehat{\mu_1'}\right]$	$\widehat{\mu_2}$ [m⁶/s²/m²]	$\widehat{\sigma}^2\left[\widehat{\mu_2}\right]$ [m¹²/s⁴/m⁴]	CoV$\left[\widehat{\mu_2}\right]$
MCS (ref.)	3000	3000	1.48×10^{-5}	3.1×10^{-14}	1%	9.3×10^{-11}	3.2×10^{-23}	6%
MCS	60	60	1.46×10^{-5}	1.2×10^{-12}	7%	7.0×10^{-11}	4.5×10^{-22}	30%
Lin	1	2.1	1.29×10^{-5}	[–]	[–]	5.5×10^{-11}	[–]	[–]
IV	1	17.0	1.55×10^{-5}	[–]	[–]	9.5×10^{-11}	[–]	[–]
CV+S+Lin	60	63.4	1.47×10^{-5}	2.7×10^{-13}	4%	8.6×10^{-11}	4.3×10^{-22}	24%
CV+S+IV	60	81.2	1.48×10^{-5}	6.1×10^{-14}	2%	9.3×10^{-11}	3.5×10^{-23}	6%

samples that is considered in order to produce *reference* results. It is observed that the seepage flow possesses a high degree of uncertainty, as its coefficient of variation is equal to $\sqrt{\widehat{\mu}_2}/\widehat{\mu}_1 = 65\%$. This is a natural consequence of the high coefficient of variation associated with the random field modeling the permeability. Moreover, it is observed that both the estimators for the mean and variance possess a good level of precision, as their coefficients of variation are relatively small (below 10%). The third row of Table 1 reports the results obtained with Monte Carlo simulation considering $n = 60$ simulations. For this case, the results reported for the second-order statistics are relatively similar to the reference ones, although it is possible to observe increased coefficient of variation of the estimators, particularly for the case of the variance of the seepage flow. The estimates for the second-order statistics obtained with classical perturbation and perturbation considering intervening variables are shown in the fourth and fifth rows, respectively. The results associated with intervening variables demand a higher number of equivalent simulations n_E when compared to those of classical perturbation. This is a natural consequence of the additional information required to produce the estimates. However, the estimates of the second-order statistics produced by perturbation with intervening variables are much more accurate than those produced with classical perturbation, highlighting its advantages.

The sixth row of Table 1 summarizes the results obtained by aggregating the results provided by Monte Carlo simulation in the third row with those of perturbation analysis and linear approximation in the fourth row by means of the framework provided by Control Variates with Splitting (CV+S+Lin). It is observed that the variance of the estimate for the mean decreases considerably: about one order of magnitude, from 1.2×10^{-12} to 2.7×10^{-13} (physical units are omitted for conciseness but can be found in the Table). Meanwhile, the accuracy of the estimated variance of the seepage flows $\widehat{\mu_2}$ improves, as the estimate of 8.6×10^{-11} is closer to the reference value than any of the estimates produced individually with Monte Carlo using $n = 60$ samples (7.0×10^{-11}) or perturbation with linearization (5.5×10^{-11}). However, the precision of the estimate does not improve that much, as the variance of the variance of the seepage flow remains almost unchanged when comparing the results produced with Control Variance with Splitting (4.3×10^{-22}) to those produced with Monte Carlo simulation considering $n = 60$ samples (4.5×10^{-22}). From the above comments, it is possible to conclude that while aggregating the results of Monte Carlo simulation with $n = 60$ simulations and perturbation with linear approximation does bring advantages, these advantages are not so pronounced. The reason for such behavior lies in the fact that a linear approximation of the seepage flow may not be that appropriate, particularly under the high level of uncertainty associated with the random field model for the permeability, which possesses a coefficient of variation of 100%.

Row seven of Table 1 presents the statistics obtained by aggregating the results from Monte Carlo simulation considering $n = 60$ samples with those of perturbation considering intervening variables under the umbrella of Control Variates with Splitting (CV+S+IV). It is interesting to note that the estimates for the mean and variance of the seepage flow are the same as those reported for the reference solution involving Monte Carlo simulation with $n = 3000$ samples, confirming the accuracy of the results obtained. Furthermore, the coefficients of variation associated with the mean and variance are almost the same as well,

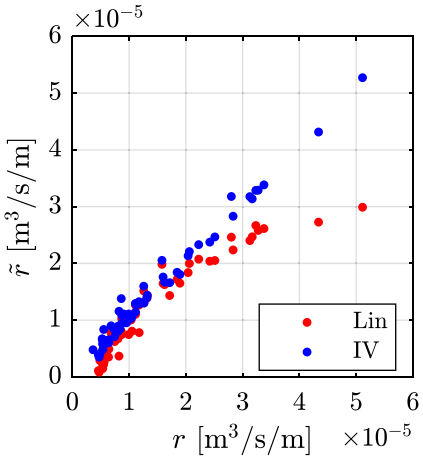

Fig. 3. Example – Comparison of seepage flow calculated using approximations ($\tilde{r}$) and the finite element model (r). Lin: linear approximation. IV: approximation considering intervening variables.

indicating that in this case, Control Variates with Splitting produces results which are as precise as those produced with Monte Carlo simulation considering $n = 3000$ simulations. As Control Variates with Splitting involves a numerical cost of about 80 equivalent simulations, there is an speedup of about $3000/80 \approx 35$ times. This highlights the fact that aggregating the results of perturbation with intervening variables and Monte Carlo simulation within the framework of Control Variates with Splitting offers a significant improvement when compared to the results produced by each approach individually. Moreover, the observed results also indicate that approximating the seepage flow considering intervening variables can perform much better than a linear approximation. In other words, perturbation considering intervening variables may offer significant advantages over classical perturbation, as discussed in detail in [29].

To gain further insight on the results discussed previously, the $n = 60$ samples associated with the sixth and seventh rows of Table 1 are plotted in Fig. 3. More specifically, the red dots in this figure illustrate pairs of seepage flows calculated for the same sample of the random field using the finite element model and the linear approximation. The blue dots are similar, except that they show the calculations with intervening variables. It is readily observed that the seepage flows predicted by the model based on intervening variables possess a higher correlation with the results provided by the finite element model than those associated with linear approximation. Such observation explains the improved estimates of the second order statistics obtained when implementing Control Variates with Splitting together with intervening variables in comparison with its counterpart considering a linear approximation. Indeed, the high correlation observed between the seepage flows calculated with the finite element model and intervening variables allows obtaining estimates with reduced variability, as discussed in Section 3.2.

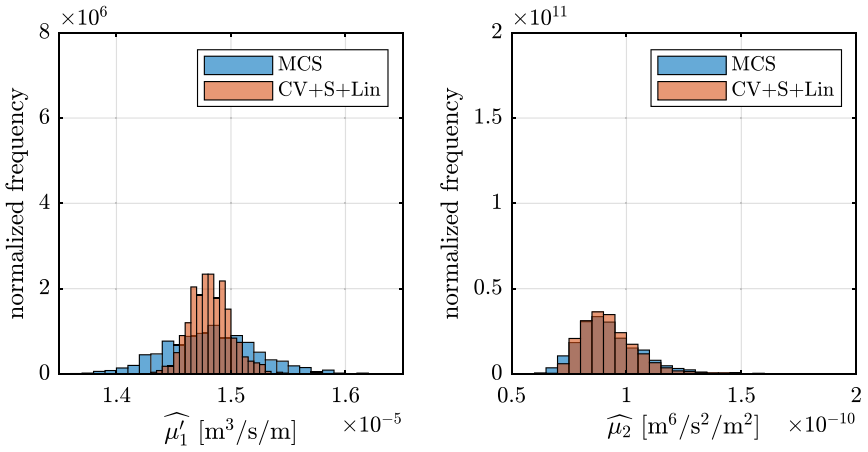

Fig. 4. Normalized histograms of the estimator of the mean $\widehat{\mu_1'}$ and variance $\widehat{\mu_2}$ considering Monte Carlo simulation (MCS) and Control Variates with Splitting involving linear approximation (CV+S+Lin) of the seepage flow.

4.3. Normalized histograms

Section 4.2 has reported simulation results obtained out of a single run of the proposed approach. In this Section, results are reported considering multiple independent runs. For this purpose, 1000 independent runs of Monte Carlo simulation are considered, each of them comprising a total of $n = 600$ simulations. In addition, 1000 independent runs of Control Variates with Splitting considering a linear approximation of the seepage flow are performed, such that the equivalent number of simulation is $n_E = 600$. Such selection for n_E ensures that the results obtained with Monte Carlo and Control Variates with Splitting are comparable between them from the point of view of numerical efforts. The above procedure produces a total of 1000 estimates of the mean and variance of the seepage flow calculated by means of Monte Carlo and Control Variates with Splitting. These estimates are used to produce *normalized* histograms of the estimated statistics, as reported in Fig. 4. In this context, normalized means that the area below the histograms is equal to 1. This provides the approximate shape of the probability density function associated with the estimators for the second order statistics. The results presented in the left hand side of Fig. 4 reveal that for the case of the mean $\widehat{\mu_1'}$, there is evident reduction in the spread of the histogram associated with Control Variates with Splitting when compared to the one produced with Monte Carlo. This is consistent with the reduced variance of the estimator for the mean observed in the sixth row of Table 1 produced with Control Variates with Splitting ($\widehat{\sigma}^2\left[\widehat{\mu_1'}\right] = 2.7 \times 10^{-13}$) when compared with the one in the third row produced with Monte Carlo ($\widehat{\sigma}^2\left[\widehat{\mu_1'}\right] = 1.2 \times 10^{-12}$). The latter is of course a qualitative comparison only, as the number of simulations involved is different. In addition, the results presented in the right hand side of Fig. 4 indicate that the histogram associated with the variance $\widehat{\mu_2}$ produced with Control Variates with Splitting presents only a slight reduction in the spread when compared to the one produced with Monte Carlo. Again, this is consistent (but not directly comparable) with the results in Table 1, as the estimate of the variance of the seepage flow with Control Variates ($\widehat{\sigma}^2\left[\widehat{\mu_1'}\right] = 4.3 \times 10^{-22}$) is slightly smaller than the one produced with Monte Carlo in the third row ($\widehat{\sigma}^2\left[\widehat{\mu_1'}\right] = 4.5 \times 10^{-22}$).

The whole procedure described above is repeated once more, except for the fact that Control Variates with Splitting is implemented considering that the approximation of the seepage flow is constructed using intervening variables. The results obtained are reported again in terms of histograms as seen in Fig. 5. In this case, it is evident that the histograms for the estimates of the mean ($\widehat{\mu_1'}$) and variance ($\widehat{\mu_2}$) of the seepage flow produced with Control Variates possess less spread than those associated with Monte Carlo. This explains the success of Control Variates in producing both accurate and precise estimates of the sought second-order statistics.

4.4. Effect of correlation length

This section investigates the effect of the correlation length on the estimates of the second-order statistics of the seepage flow. For this purpose, two extreme cases are considered, namely correlation lengths equal to $L = 2.5$ [m] and $L = 100$ [m]. The former corresponds to a weakly correlated case while the latter is a strongly correlated case, such that the random field model reduces to a single random variable. For the weakly correlated case, it is necessary to consider $n_\theta = 109$ terms to retain 95% of the total variability in the random field representation while for the strongly correlated case, $n_\theta = 1$ suffices.

Table 2 presents the estimates of the second order statistics of the seepage flow for the aforementioned correlation lengths as well as for the correlation length of $L = 10$ [m] considered in the original formulation of the problem. All the results were calculated by aggregating the approximate solution involving perturbation with intervening variables and $n = 60$ Monte Carlo samples of the exact response. The results obtained indicate that both the mean and variance of the seepage flow tend to increase as the correlation length increases. For a short correlation length, it is expected that the permeability varies considerably with distance. And as the seepage flow captures an average behavior, the effects of spatial variability tend to cancel out. The effect is exactly the opposite for a long correlation length.

From the point of view of the coefficient of variation of the estimators reported in Table 2, it is interesting to note that the coefficient associated with the variance (last column of Table 2) decreases as the correlation length increases. To understand this behavior, Fig. 6 compares the $n = 60$ exact and approximate seepage flow responses employed for the calculation with Control Variates for the two correlation lengths considered in this Section. The plot on the left shows a high degree of correlation between the exact and approximate responses while the plot on the right shows an almost perfect correlation. The latter effect is due to the fact that the random field model reduces to a single random variable. In such case, the approximate response based on perturbation considering intervening variables is capable of emulating almost exactly the exact response. Such an effect had already been observed in the past when applying intervening variables, see e.g. [29]. As such, for the strongly correlated case, the approximate model is practically as good as the exact one and hence, it is capable of producing very accurate estimates, explaining the small coefficients of variation of the estimators observed in Table 2 for $L = 100$ [m] for both mean and variance of the seepage flow (last row of Table 2).

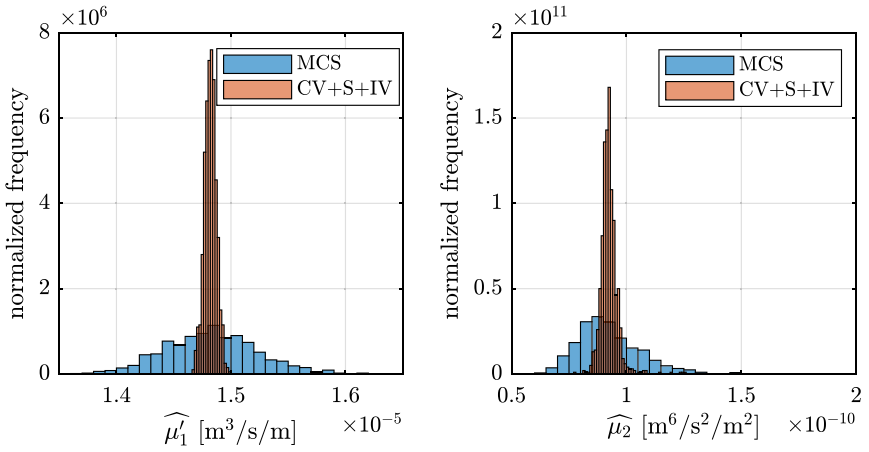

Fig. 5. Normalized histograms of the estimator of the mean $\widehat{\mu_1'}$ and variance $\widehat{\mu_2}$ considering Monte Carlo simulation (MCS) and Control Variates with Splitting involving intervening variables (CV+S+IV) for approximation of the seepage flow.

Table 2
Second-order statistics of seepage flow as a function of correlation length L.

L [m]	$\widehat{\mu_1'}$ [m³/s/m]	$\widehat{\sigma}^2\left[\widehat{\mu_1'}\right]$ [m⁶/s²/m²]	CoV$\left[\widehat{\mu_1'}\right]$	$\widehat{\mu_2}$ [m⁶/s²/m²]	$\widehat{\sigma}^2\left[\widehat{\mu_2}\right]$ [m¹²/s⁴/m⁴]	CoV$\left[\widehat{\mu_2}\right]$
2.5	1.31×10^{-5}	8.4×10^{-15}	0.7%	9.1×10^{-12}	6.6×10^{-25}	9%
10	1.48×10^{-5}	6.1×10^{-14}	2%	9.3×10^{-11}	3.5×10^{-23}	6%
100	1.82×10^{-5}	7.1×10^{-20}	< 0.1%	3.3×10^{-10}	9.2×10^{-27}	< 0.1%

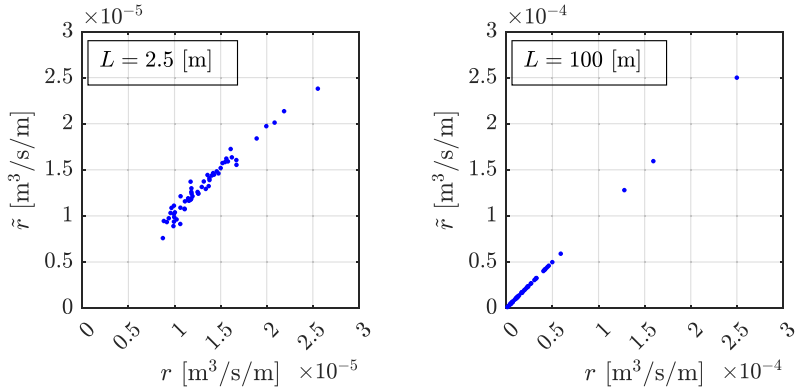

Fig. 6. Example – Comparison of seepage flow calculated using intervening variables ($\tilde{r}$) and the finite element model (r) for different values of correlation length L.

5. Discussion and conclusions

This contribution has explored the application of Control Variates with Splitting for aggregating the results produced with Monte Carlo simulation and perturbation analysis. The scope of application considers the estimation of second-order statistics of the response of stochastic finite element models of linear systems under steady-state conditions. The results obtained indicate that:

- Control Variates offers a framework for aggregating results such that both accuracy and precision of the sought statistics can be improved with respect to those statistics estimated with each approach separately. In the application example, such behavior was clearly observed for the case where Monte Carlo was aggregated with perturbation analysis involving intervening variables.
- The success of aggregating results is directly related with the ability of the approximate response (associated with perturbation analysis) for mimicking the exact response, as this will increase the covariance between the estimators of the approximate and exact models. In the worst-case scenario where the approximate response does not exhibit significant covariance with the exact model, Control Variates with Splitting retains by construction the results provided by Monte Carlo simulation and gives less weight to the results produced with perturbation analysis.
- Numerical efforts associated with the implementation of Control Variates with Splitting are not significant. In fact, no additional evaluations of the system's response are required. It is only required to keep track of the sample subsets involved in the Splitting technique.

There are several open paths for further development of the framework involving Control Variates and Splitting. One of them would be constructing the approximation for the system's response using concepts of active learning. In that way, the system's response can be approximated on the fly, allocating numerical resources spent in evaluating the exact response in an optimal way. Another path for development is the application of Control Variates with Splitting to problems involving a large number of random variables, possibly in the order of hundreds. In principle, such extension should be possible as long as there is a high correlation between the exact and approximate models. The latter assertion can be easily verified by reviewing the expressions for calculating second-order statistics by means of Control Variates with Splitting, that depend explicitly on correlation between models but not on the dimensionality of the associated random variable vector.

An additional path for extending the range of application of Control Variates with Splitting is investigating the advantages of aggregating methods other than a perturbation approach. In this sense, it should be noted that, in principle, the framework provided by Control Variates is agnostic with respect to the specific approach considered to approximate the system's response.

CRediT authorship contribution statement

Cristóbal H. Acevedo: Writing – review & editing, Software, Methodology, Formal analysis, Conceptualization. **Marcos A. Valdebenito:** Writing – review & editing, Writing – original draft, Supervision, Methodology, Funding acquisition, Formal analysis, Conceptualization. **Iván V. González:** Writing – review & editing, Methodology, Formal analysis, Conceptualization. **Héctor A. Jensen:** Writing – review & editing, Methodology, Formal analysis, Conceptualization. **Matthias G.R. Faes:** Writing – review & editing, Methodology, Formal analysis, Conceptualization. **Yong Liu:** Writing – review & editing, Methodology, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

This research is partially supported by ANID (National Agency for Research and Development, Chile) under grant number 1200087 and its graduate scholarship program CONICYT-PFCHA/MagisterNacional/2019-22191834 and CONICYT-PFCHA/MagisterNacional/2017-22171799. This support is gratefully acknowledged by the authors.

Appendix A. Accuracy and precision

The terms *accuracy* and *precision* are often used to characterize the appropriateness of measurements. As pointed out in [64], accuracy reflects the "closeness of agreement between a quantity value obtained by measurement and the true value of the measurand" while precision represents "the closeness of agreement between independent test results obtained under stipulated conditions". Within the context of this work, accuracy and precision are particularly suitable to characterize the estimates of second-order statistics obtained using perturbation analysis and Monte Carlo simulation. In perturbation analysis, it is expected that the estimates for second-order statistics are precise (because they are a deterministic value) but possibly not accurate (because the response is approximated). If the estimator were to be treated as a probability distribution, it would actually correspond to a Dirac delta, as the estimated statistics are precise numbers. This situation is illustrated schematically in Fig. A.7, where the aforementioned Dirac delta is denoted with a green arrow. As noted from the Figure, there may be differences between the true statistic and the prediction by perturbation analysis, which corresponds to lack of accuracy. On the contrary, an estimator produced with Monte Carlo simulation will be accurate, because it is calculated based on the actual response of the system. This is reflected in Fig. A.7 with the red and blue curves denoting probability distribution of estimators which were calculated with different number of samples, where the blue curve has associated a number of samples larger than that of the red curve. The expected value of both probability distributions lies at the true statistic. However,

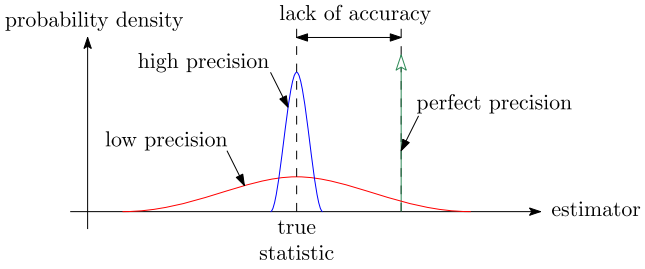

Fig. A.7. Schematic representation of the concepts of accuracy and precision.

the probability distribution associated with the blue curve offers an estimate of the statistic with higher precision than that associated with the red one, as it possesses less dispersion. The latter is a direct consequence of the number of samples considered for estimating the statistics, which is larger for the blue curve.

Appendix B. Calculation of exponent associated with exponential intervening variables

The criterion for calculating the exponent $m_k,\ k = 1, \dots, n_\theta$ associated with the exponential intervening variable consists of imposing the condition that the second order partial derivatives of the system's response match the second order derivatives of the approximate system's response. This last condition is expressed as:

$$\left.\frac{\partial^2 r}{\partial \theta_k^2}\right|_{\theta=\theta^{(0)}} = \left.\frac{\partial^2 r^1}{\partial \theta_k^2}\right|_{\theta=\theta^{(0)}},\ k = 1, \dots, n_\theta. \tag{B.1}$$

Note that for applying the above equation, knowledge on the diagonal terms of the Hessian matrix of the system's response is required. This is quite advantageous for high-dimensional problems, as off-diagonal terms are not required. Thus, enforcing the condition of Eq. (B.1), the coefficients $m_k,\ k = 1, \dots, n_\theta$ are calculated by means of the following formula:

$$m_k = \frac{r_{,kk}}{r_{,k}},\ k = 1, \dots, n_\theta. \tag{B.2}$$

In the above equation, $r_{,kk} = \partial^2 r/\partial \theta_k^2\big|_{\theta=\theta^{(0)}}$. Note that the definition for the coefficient m_k in Eq. (B.2) is undefined in case the associated first-order derivative is equal to zero and, moreover, may grow rapidly for small values of $r_{,k}$. These issues are prevented if the following expression is considered for calculating m_k:

$$m_k = \begin{cases} 1 & \text{if } r_{,k} = 0 \\ m_k^\star & \text{if } r_{,k} \neq 0 \end{cases} \tag{B.3}$$

where $m_k^\star$ is defined as:

$$m_k^\star = \begin{cases} -3\chi_k & \text{if } \frac{r_{,kk}}{r_{,k}} \leq -3\chi_k \\ \frac{r_{,kk}}{r_{,k}} & \text{if } -3\chi_k \leq \frac{r_{,kk}}{r_{,k}} \leq 3\chi_k \\ 3\chi_k & \text{if } \frac{r_{,kk}}{r_{,k}} \geq 3\chi_k \end{cases} \tag{B.4}$$

where χ_k is a coefficient measuring the mean variability associated with the kth term of the underlying Gaussian field involved in the representation of the random field [29].

According to Eq. (B.3), whenever the associated first-order derivative is equal to zero, m_k is set arbitrarily as 1. Such arbitrary value is irrelevant, as the first-order derivative is equal to zero and does not affect the first order Taylor expansion. Moreover, the criterion in Eq. (B.4) prevents the coefficient m_k from growing unboundedly (see, e.g. [56]). In particular, the bound for the magnitude of the coefficient m_k is chosen as $3\chi_k$, following the recommendations reported in [29,65]. Numerical validations as carried out in [29] suggest that such selection is appropriate.

Appendix C. Estimation of central co-moments

The bivariate central co-moment between r and $\tilde{r}$ of order (p,q) is:

$$\mu_{p,q}(r,\tilde{r}) = \mathbb{E}\left[(r-\mathbb{E}[r])^p(\tilde{r}-\mathbb{E}[\tilde{r}])^q\right] \tag{C.1}$$

where $\mathbb{E}[\cdot]$ denotes expected value. Unbiased estimators $\widehat{\mu_{p,q}}$ for this bivariate co-moment can be calculated by means of Monte Carlo simulation [66] considering a set of l samples of the uncertain input variable vector, collected in matrix $\boldsymbol{\Theta}_l$, as shown in the equations below for different values of p and q. Note that in these equations, the auxiliary variable $s_{p,q}$ is introduced such that $s_{p,q} = \sum_{i=1}^{l}\left(r\left(\boldsymbol{\theta}^{(i)}\right)\right)^p\left(\tilde{r}\left(\boldsymbol{\theta}^{(i)}\right)\right)^q$, where $\boldsymbol{\theta}^{(i)}$ is the ith sample of the sample set $\boldsymbol{\Theta}_l$.

$$\widehat{\mu_{2,0}} = \frac{l s_{2,0} - s_{1,0}^2}{(l-1)l} \tag{C.2}$$

$$\widehat{\mu_{1,1}} = \frac{l s_{1,1} - s_{0,1}s_{1,0}}{(l-1)l} \tag{C.3}$$

$$\widehat{\mu_{0,2}} = \frac{l s_{0,2} - s_{0,1}^2}{(l-1)l} \tag{C.4}$$

$$\begin{aligned}\widehat{\mu_{2,2}} = \frac{1}{(l-3)(l-2)(l-1)l}\Big(&\left(-2l^2+4l-6\right)s_{2,1}s_{0,1} + \left(-2l^2+4l-6\right)s_{1,0}s_{1,2} + \\ &\left(l^3-2l^2+3l\right)s_{2,2} + l s_{2,0}s_{0,1}^2 + 4l s_{1,0}s_{1,1}s_{0,1} + l s_{0,2}s_{1,0}^2 + \\ &(6-4l)s_{1,1}^2 + (3-2l)s_{0,2}s_{2,0} - 3s_{1,0}^2s_{0,1}^2\Big)\end{aligned} \tag{C.5}$$

$$\begin{aligned}\widehat{\mu_{4,0}} = \frac{1}{(l-3)(l-2)(l-1)l}\Big(&\left(-4l^2+8l-12\right)s_{3,0}s_{1,0} + \left(l^3-2l^2+3l\right)s_{4,0} + \\ &6l s_{2,0}s_{1,0}^2 + (9-6l)s_{2,0}^2 - 3s_{1,0}^4\Big)\end{aligned} \tag{C.6}$$

$$\begin{aligned}\widehat{\mu_{0,4}} = \frac{1}{(l-3)(l-2)(l-1)l}\Big(&\left(-4l^2+8l-12\right)s_{0,3}s_{0,1} + \left(l^3-2l^2+3l\right)s_{0,4} + \\ &6l s_{0,2}s_{0,1}^2 + (9-6l)s_{0,2}^2 - 3s_{0,1}^4\Big)\end{aligned} \tag{C.7}$$

The unbiased estimators for the squared co-moments $\mu_{1,1}^2$, $\mu_{2,0}^2$, $\mu_{0,2}^2$ and the co-moment product $\mu_{2,0}\mu_{0,2}$ are the following [66].

$$\begin{aligned}\widehat{\mu_{1,1}^2} = \frac{1}{(l-3)(l-2)(l-1)l}\Big(&\left(l^2-3l+2\right)s_{1,1}^2 + \left(l-l^2\right)s_{2,2} + \\ &(2-2l)s_{1,0}s_{1,1}s_{0,1} + (2l-2)s_{2,1}s_{0,1} + (2l-2)s_{1,0}s_{1,2} + \\ &s_{1,0}^2s_{0,1}^2 - s_{2,0}s_{0,1}^2 - s_{0,2}s_{1,0}^2 + s_{0,2}s_{2,0}\Big)\end{aligned} \tag{C.8}$$

$$\widehat{\mu_{2,0}^2} = \frac{\left(l^2-3l+3\right)s_{2,0}^2 + \left(l-l^2\right)s_{4,0} - 2l s_{2,0}s_{1,0}^2 + (4l-4)s_{3,0}s_{1,0} + s_{1,0}^4}{(l-3)(l-2)(l-1)l} \tag{C.9}$$

$$\widehat{\mu_{0,2}^2} = \frac{\left(l^2-3l+3\right)s_{0,2}^2 + \left(l-l^2\right)s_{0,4} - 2l s_{0,2}s_{0,1}^2 + (4l-4)s_{0,3}s_{0,1} + s_{0,1}^4}{(l-3)(l-2)(l-1)l} \tag{C.10}$$

$$\begin{aligned}\widehat{\mu_{2,0}\mu_{0,2}} = \frac{1}{(l-3)(l-2)(l-1)l}\Big(&\left(l^2-3l+1\right)s_{0,2}s_{2,0} + \left(l-l^2\right)s_{2,2} + \\ &(2-l)s_{2,0}s_{0,1}^2 + (2l-2)s_{2,1}s_{0,1} + (2-l)s_{0,2}s_{1,0}^2 + \\ &(2l-2)s_{1,0}s_{1,2} + s_{1,0}^2s_{0,1}^2 - 4s_{1,0}s_{1,1}s_{0,1} + 2s_{1,1}^2\Big)\end{aligned} \tag{C.11}$$

References

[1] Ghanem R, Spanos P. Polynomial chaos in stochastic finite elements. J Appl Mech (ASME) 1990;57(1):197–202.

[2] Jensen H, Iwan W. Response of systems with uncertain parameters to stochastic excitation. J Eng Mech 1992;118(5):1012–25.

[3] Ávila da Silva Jr C, Beck A. A fast convergence parameter for Monte Carlo-Neumann solution of linear stochastic systems. ASCE-ASME J Risk Uncertain Eng Syst B Mech Eng 2015;1(2):021002. http://dx.doi.org/10.1115/1.4029741, arXiv:https://asmedigitalcollection.asme.org/risk/article-pdf/1/2/021002/6070906/risk_1_2_021002.pdf.

[4] Schuëller G. On the treatment of uncertainties in structural mechanics and analysis. Comput Struct 2007;85(5–6):235–43. http://dx.doi.org/10.1016/j.compstruc.2006.10.009.

[5] Faes M, Moens D. Recent trends in the modeling and quantification of non-probabilistic uncertainty. Arch Comput Methods Eng 2020;27(3):633–71. http://dx.doi.org/10.1007/s11831-019-09327-x.

[6] Fina M, Panther L, Weber P, Wagner W. Shell buckling with polymorphic uncertain surface imperfections and sensitivity analysis. ASCE-ASME J Risk Uncertain Eng Syst B Mech Eng 2021;7(2).

[7] Der Kiureghian A, Ke J-B. The stochastic finite element method in structural reliability. Probab Eng Mech 1988;3(2):83–91.

[8] Fang P-P, Liu Y, Shields M. Direct simulation methods for a class of normal and lognormal random fields with applications in modeling material properties. J Eng Mech 2021;148(2):04021146. http://dx.doi.org/10.1061/(asce)em.1943-7889.0002076.

[9] Ou-Yang J-Y, Li D-Q, Tang X-S, Liu Y. A patching algorithm for conditional random fields in modeling material properties. Comput Methods Appl Mech Engrg 2021;377:113719. http://dx.doi.org/10.1016/j.cma.2021.113719, https://www.sciencedirect.com/science/article/pii/S0045782521000554.

[10] Hu Z, Du X. A random field approach to reliability analysis with random and interval variables. ASCE-ASME J Risk Uncertain Eng Syst B Mech Eng 2015;1(4):041005. http://dx.doi.org/10.1115/1.4030437, arXiv:https://asmedigitalcollection.asme.org/risk/article-pdf/1/4/041005/6071907/risk_1_4_041005.pdf.

[11] Shinozuka M. Structural response variability. J Eng Mech 1987;113(6):825–42. http://dx.doi.org/10.1061/(ASCE)0733-9399(1987)113:6(825).

[12] Papadopoulos V, Deodatis G, Papadrakakis M. Flexibility-based upper bounds on the response variability of simple beams. Comput Methods Appl Mech Engrg 2005;194(12):1385–404. http://dx.doi.org/10.1016/j.cma.2004.06.040, special Issue on Computational Methods in Stochastic Mechanics and Reliability Analysis, http://www.sciencedirect.com/science/article/pii/S0045782504003986.

[13] Shinozuka M, Jan C-M. Digital simulation of random processes and its applications. J Sound Vib 1972;25(1):111–28. http://dx.doi.org/10.1016/0022-460X(72)90600-1, https://www.sciencedirect.com/science/article/pii/0022460X72906001.

[14] Melchers R. Importance sampling in structural systems. Struct Saf 1989;6(1):3–10.

[15] Au S, Beck J. Estimation of small failure probabilities in high dimensions by subset simulation. Probab Eng Mech 2001;16(4):263–77.

[16] Vanmarcke E, Shinozuka M, Nakagiri S, Schuëller G, Grigoriu M. Random fields and stochastic finite elements. Struct Saf 1986;3(3):143–66. http://dx.doi.org/10.1016/0167-4730(86)90002-0, https://www.sciencedirect.com/science/article/pii/0167473086900020.

[17] Yamazaki F, Shinozuka M, Dasgupta G. Neumann expansion for stochastic finite element analysis. J Eng Mech 1988;114(8):1335–54. http://dx.doi.org/10.1061/(ASCE)0733-9399(1988)114:8(1335).

[18] Kamiński M. On generalized stochastic perturbation-based finite element method. Commun Numer Methods Eng 2006;22(1):23–31.

[19] Blatman G, Sudret B. An adaptive algorithm to build up sparse polynomial chaos expansions for stochastic finite element analysis. Probab Eng Mech 2010;25(2):183–97. http://dx.doi.org/10.1016/j.probengmech.2009.10.003, http://www.sciencedirect.com/science/article/pii/S0266892009000666.

[20] Deodatis G, Shinozuka M. Bounds on response variability of stochastic systems. J Eng Mech 1989;115(11):2543–63. http://dx.doi.org/10.1061/(ASCE)0733-9399(1989)115:11(2543).

[21] Papadopoulos V, Kalogeris I. A Galerkin-based formulation of the probability density evolution method for general stochastic finite element systems. Comput Mech 2016;57(5):701–16. http://dx.doi.org/10.1007/s00466-015-1256-9.

[22] Huo H, Xu W, Wang W, Chen G, Yang D. New non-intrusive stochastic finite element method for plate structures. Comput Struct 2022;268:106812. http://dx.doi.org/10.1016/j.compstruc.2022.106812, https://www.sciencedirect.com/science/article/pii/S0045794922000724.

[23] Chen G, Yang D. Direct probability integral method for stochastic response analysis of static and dynamic structural systems. Comput Methods Appl Mech Engrg 2019;357:112612. http://dx.doi.org/10.1016/j.cma.2019.112612, http://www.sciencedirect.com/science/article/pii/S0045782519304888.

[24] Li J, Chen J, Sun W, Peng Y. Advances of the probability density evolution method for nonlinear stochastic systems. Probab Eng Mech 2012;28:132–42. http://dx.doi.org/10.1016/j.probengmech.2011.08.019, http://www.sciencedirect.com/science/article/pii/S0266892011000749.

[25] Elishakoff I, Ren Y, Shinozuka M. Improved finite element method for stochastic problems. Chaos Solitons Fractals 1995;5(5):833–46. http://dx.doi.org/10.1016/0960-0779(94)00157-L, http://www.sciencedirect.com/science/article/pii/096007799400157L.

[26] Ghanem R, Spanos P. Stochastic finite elements: A spectral approach. New York: Springer; 1991.

[27] Panayirci H, Pradlwarter H, Schuëller G. Efficient stochastic structural analysis using Guyan reduction. Adv Eng Softw 2011;42(4):187–96.

[28] Fuchs M, Shabtay E. The reciprocal approximation in stochastic analysis of structures. Chaos Solitons Fractals 2000;11(6):889–900.

[29] Valdebenito M, Labarca A, Jensen H. On the application of intervening variables for stochastic finite element analysis. Comput Struct 2013;126:164–76. http://dx.doi.org/10.1016/j.compstruc.2013.01.001.

[30] Fishman G. Monte Carlo: concepts, algorithms and applications. New York, NY: Springer; 1996.

[31] Charlton T, Rouainia M, Dawson R. Control variate approach for efficient stochastic finite-element analysis of geotechnical problems. ASCE-ASME J Risk Uncertain Eng Syst A 2018;4(3):04018031. http://dx.doi.org/10.1061/AJRUA6.0000983, arXiv:https://ascelibrary.org/doi/pdf/10.1061/AJRUA6.0000983, https://ascelibrary.org/doi/abs/10.1061/AJRUA6.0000983.

[32] Nelson B. Control variate remedies. Oper Res 1990;38(6):974–92. http://dx.doi.org/10.1287/opre.38.6.974.

[33] Pasupathy R, Schmeiser B, Taaffe M, Wang J. Control-variate estimation using estimated control means. IIE Trans 2012;44(5):381–5. http://dx.doi.org/10.1080/0740817X.2011.610430.

[34] Rashki M, Ghavidel A, Ghohani Arab H, Mousavi S. Low-cost finite element method-based reliability analysis using adjusted control variate technique. Struct Saf 2018;75:133–42, http://www.sciencedirect.com/science/article/pii/S0167473017301170.

[35] Boyaval S. A fast Monte-Carlo method with a reduced basis of control variates applied to uncertainty propagation and Bayesian estimation. Comput Methods Appl Mech Eng 2012;241–244:190–205. http://dx.doi.org/10.1016/j.cma.2012.05.003.

[36] Geraci G, Eldred M, Iaccarino G. A multifidelity control variate approach for the multilevel monte carlo technique. Center for Turbulence Research, Annual Research Brief; 2015.

[37] Giles M. Multilevel Monte Carlo path simulation. Oper Res 2008;56(3):607–17. http://dx.doi.org/10.1287/opre.1070.0496.

[38] Gorodetsky A, Geraci G, Eldred M, Jakeman J. A generalized approximate control variate framework for multifidelity uncertainty quantification. J Comput Phys 2020;408:109257. http://dx.doi.org/10.1016/j.jcp.2020.109257, https://www.sciencedirect.com/science/article/pii/S0021999120300310.

[39] Pisaroni M, Nobile F, Leyland P. A continuation multi level monte carlo (C-MLMC) method for uncertainty quantification in compressible inviscid aerodynamics. Comput Methods Appl Mech Engrg 2017;326:20–50. http://dx.doi.org/10.1016/j.cma.2017.07.030, https://www.sciencedirect.com/science/article/pii/S0045782516308155.

[40] Avramidis A, Wilson J. A splitting scheme for control variates. Oper Res Lett 1993;14(4):187–98. http://dx.doi.org/10.1016/0167-6377(93)90069-S, http://www.sciencedirect.com/science/article/pii/016763779390069S.

[41] Vidal-Codina F, Nguyen N, Giles M, Peraire J. A model and variance reduction method for computing statistical outputs of stochastic elliptic partial differential equations. J Comput Phys 2015;297:700–20. http://dx.doi.org/10.1016/j.jcp.2015.05.041, http://www.sciencedirect.com/science/article/pii/S0021999115003757.

[42] Bathe K. Finite element procedures. New Jersey: Prentice Hall; 1996.

[43] Stefanou G. The stochastic finite element method: Past, present and future. Comput Methods Appl Mech Engrg 2009;198(9–12):1031–51.

[44] Liu P, Der Kiureghian A. Multivariate distribution models with prescribed marginals and covariances. Probab Eng Mech 1986;1(2):105–12.

[45] Grigoriu M. Simulation of stationary non-Gaussian translation processes. J Eng Mech 1998;124(2):121–6. http://dx.doi.org/10.1061/(ASCE)0733-9399(1998)124:2(121).

[46] Faes M, Broggi M, Chen G, Phoon K-K, Beer M. Distribution-free P-box processes based on translation theory: Definition and simulation. Probab Eng Mech 2022;69:103287. http://dx.doi.org/10.1016/j.probengmech.2022.103287, https://www.sciencedirect.com/science/article/pii/S0266892022000558.

[47] Ayyub B, McCuen R. Probability, statistics, and reliability for engineers and scientists. 3rd ed.. Taylor & Francis; 2011, https://books.google.de/books?id=zfnltuHc5Y4C.

[48] Ang A, Tang W. Probability concepts in engineering: emphasis on applications to civil and environmental engineering. Wiley; 2007.

[49] Liu W, Mani A, Belytschko T. Finite element methods in probabilistic mechanics. Probab Eng Mech 1987;2(4):201–13.

[50] Schuëller G, editor. A state-of-the-art report on computational stochastic mechanics. Probab Eng Mech 1997;12(4):197–321.

[51] Rahman S, Rao B. A perturbation method for stochastic meshless analysis in elastostatics. Internat J Numer Methods Engrg 2001;50(8):1969–91.

[52] Haftka R, Gürdal Z. Elements of structural optimization. 3rd ed.. Dordrecht, The Netherlands: Kluwer; 1992.

[53] Matthies H, Brenner C, Bucher C, Soares C. Uncertainties in probabilistic numerical analysis of structures and solids - stochastic finite elements. Struct Saf 1997;19(3):283–336.

[54] Schmit L, Farshi B. Some approximation concepts for structural synthesis. AIAA J 1974;12(5):692–9.

[55] Prasad B. Explicit constraint approximation forms in structural optimization. Part 1: Analyses and projections. Comput Methods Appl Mech Eng 1983;40(1):1–26. http://dx.doi.org/10.1016/0045-7825(83)90044-0.

[56] Fadel G, Riley M, Barthelemy J. Two point exponential approximation method for structural optimization. Struct Optim 1990;2(2):117–24.

[57] Nelson B. Batch size effects on the efficiency of control variates in simulation. European J Oper Res 1989;43(2):184–96. http://dx.doi.org/10.1016/0377-2217(89)90212-9, http://www.sciencedirect.com/science/article/pii/0377221789902129.

[58] Box M. Bias in nonlinear estimation. J R Stat Soc Ser B (Methodological) 1971;33(2):171–201, http://www.jstor.org/stable/2985002.

[59] Griffiths D, Fenton G. Seepage beneath water retaining structures founded on spatially random soil. Geotechnique 1993;43(4):577–87.

[60] Spanos P, Beer M, Red-Horse J. Karhunen-loéve expansion of stochastic processes with a modified exponential covariance kernel. J Eng Mech 2007;133(7):773–9. http://dx.doi.org/10.1061/(ASCE)0733-9399(2007)133:7(773).

[61] Faes M, Broggi M, Spanos P, Beer M. Elucidating appealing features of differentiable auto-correlation functions: A study on the modified exponential kernel. Probab Eng Mech 2022;69:103269. http://dx.doi.org/10.1016/j.probengmech.2022.103269, https://www.sciencedirect.com/science/article/pii/S026689202200042X.

[62] Liu Y, Wang M-Y, Pan Y-T, Yao K. Large-scale 3D random finite element analysis of embankment seepage stability. In: Weng M-C, Kawamura S, Ding J, editors. Advancements in geotechnical engineering. Cham: Springer International Publishing; 2021, p. 1–13.

[63] Cho S. First-order reliability analysis of slope considering multiple failure modes. Eng Geol 2013;154:98–105, https://www.sciencedirect.com/science/article/pii/S0013795213000045.

[64] Menditto A, Patriarca M, Magnusson B. Understanding the meaning of accuracy, trueness and precision. Accreditation Qual Assur 2007;12(1):45–7. http://dx.doi.org/10.1007/s00769-006-0191-z.

[65] Wood D, Groenwold A. Non-convex dual forms based on exponential intervening variables, with application to weight minimization. Internat J Numer Methods Engrg 2009;80(12):1544–72.

[66] Stokes B. mathStatica 2.5. J Stat Softw Softw Rev 2012;47(1):1–12, https://www.jstatsoft.org/v047/s01.