Computers and Structures 321 (2026) 108054

Contents lists available at ScienceDirect

Computers and Structures

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

# Scale-invariant Monte Carlo and multilevel Monte Carlo estimation of mean and variance: An application to simulation of linear elastic bone tissue

Sharana Kumar Shivanand [a,*], Bojana Rosić [b]

[a] Institute of Scientific Computing, Technische Universität Braunschweig, Braunschweig, Germany
[b] Applied Mechanics and Data Analysis, University of Twente, Enschede, Netherlands

ARTICLE INFO

*Keywords:*
Monte Carlo
Multilevel Monte Carlo
Normalized error
Uncertainty quantification
Linear elasticity
h-statistics
Random anisotropy
Bone tissue

ABSTRACT

We propose novel scale-invariant error estimators for the Monte Carlo and multilevel Monte Carlo estimation of mean and variance. For any linear transformation of the distribution of the quantity of interest, the computation cost across fidelity levels is optimized using a normalized error estimate, which is not only fully dimensionless but also remains robust to variations in the characteristics of the distribution. We demonstrate the effectiveness of the algorithms through application to a mechanical simulation of linear elastic bone tissue, where material uncertainty incorporating both heterogeneity and random anisotropy is considered in the constitutive law.

## 1. Introduction

The estimation of statistical moments for characterizing a probabilistic quantity of interest (QoI), such as the solution to a stochastic partial differential equation (SPDE) or an ordinary differential equation (SODE), is of fundamental importance in the field of uncertainty quantification (UQ) [38,67]. Specifically, let $X$ be a random variable defined on a probability space $(\Omega, \mathfrak{F}, \mathbb{P})$, where $\Omega$ denotes the sample space, $\mathfrak{F}$ is a $\sigma$-algebra of measurable events, and $\mathbb{P}$ is the associated probability measure. The $p$-th central moment of $X$, for $p \in \mathbb{N}$, is defined as

$$\mu_p(X) = \mathbb{E}\left[(X - \mathbb{E}(X))^p\right], \tag{1}$$

where, $\mathbb{E}(X) := \mu(X)$ is the mean or the first raw moment (not to be confused with the first central moment $\mu_1$, which equals zero when $p = 1$). The second central moment, when $p = 2$, $\mu_2(X) = \mathbb{V}\mathrm{ar}(X) := \mathbb{E}\left((X - \mathbb{E}(X))^2\right)$, denotes the variance. In contrast, the third and fourth central moments are typically expressed in standardized form: the skewness and kurtosis, which are defined as $\alpha_3 = \mu_3/\sqrt{\mathbb{V}\mathrm{ar}^3}$ and $\alpha_4 = \mu_4/\mathbb{V}\mathrm{ar}^2$, respectively. While skewness and kurtosis are dimensionless and invariant under linear transformation of $X$, the mean and variance are inherently scale-dependent. Consequently, their estimation accuracy, typically evaluated using absolute error metrics such as the mean squared error (MSE) or root mean squared error (RMSE), also depends on the scale of the QoI. This scale dependency presents an interpretability challenge, especially in practical scenarios where comparisons across different estimators or between the same estimator applied to differently scaled QoI are required. Addressing this, normalized error estimates play a crucial role in statistics by providing a standardized measure of the accuracy of statistical estimates.

Normalization in statistics is a broad topic with multiple interpretations [32]. One common usage refers to the standardization of observational data (or the QoI), often represented by the z-score or standard score, defined as $z = X - \mu/\sqrt{\mathbb{V}\mathrm{ar}}$. Another interpretation involves the normalization of statistical moments of the form, $\alpha_p = \mathbb{E}(z^p) = \mu_p/\sqrt{\mathbb{V}\mathrm{ar}^p}$, whereby the first and second standardized moments, $\alpha_1$ and $\alpha_2$— representing the mean and variance of the standardized variable $z$—are fixed at 0 and 1, respectively. This implies that these moments are invariant across different distributions, offering no distinctive information. In contrast, higher-order standardized moments, such as skewness $\alpha_3$ and kurtosis $\alpha_4$, retain distribution-specific characteristics and are thus useful for distinguishing among probability distributions. Therefore, in this work, the primary focus is on the estimation of the mean $\mu$ and variance $\mu_2$, where normalization is applied not to the QoI or the moments, but rather to their scale-dependent absolute error estimates.

In this study, we focus on random sampling-based statistical estimators, such as the Monte Carlo (MC) method; traditionally regarded as the gold standard for solving stochastic problems, due to its simplicity and resilience against the curse of dimensionality [16,22,27,47].

* Corresponding author.
*Email address:* sharana.shivanand@kit.edu (S.K. Shivanand).

https://doi.org/10.1016/j.compstruc.2025.108054
Received 22 July 2025; Accepted 24 November 2025
Available online 6 December 2025

However, despite its appeal, the practical implementation of MC often faces challenges due to slow convergence, thereby demanding substantial computational effort. To overcome these limitations, recent research has increasingly turned to variance reduction techniques, notably the multilevel Monte Carlo (MLMC) method. A key objective of the MLMC method is to spread out the sampling strategy across a hierarchy of different fidelities (or levels) such that the number of stochastic samples drastically decreases with the increment in fidelity of the model. Under the right conditions, this results in an overall reduction of the computational cost as compared to the MC approach.

To the best of the authors' knowledge, MLMC was first introduced by [26] in the context of estimating multi-dimensional parameter-dependent integrals and was further employed by [20] for solving Itô's stochastic ordinary differential equations in the computational finance applications. Following this, MLMC was extended to solve the linear elliptic PDEs describing the subsurface flow with inhomogeneous stochastic parameters [2,9]. A further analysis of modelling of rough random field coefficients of elliptic PDEs on MLMC convergence was studied in [7,65]. However, the previously mentioned works focused only on the approximation of the QoI's sample mean; therefore, they lacked a full characterization of the probabilistic solution. Consequently, [48] studied the unbiased MLMC sample variance estimator, which was further analysed in [4] on a class of elliptic random obstacle problems—nonetheless, the corresponding error estimates were defined only on worst case bounds. To address this limitation, an alternative MLMC variance estimator based on h-statistics [14,51] was introduced in [35], which emphasizes on unbiased construction of the MSE in closed form, an approach also adopted in this study. More recently, [46] developed MLMC estimators for the biased standard deviation and its linear combination with the unbiased mean, which are important for optimization under uncertainty (OUU) workflows. Though the estimation of other higher-order moments such as skewness and kurtosis or covariance structures of the QoI via MLMC is out of the scope of this paper, more literature on this can be found in [5,35,57].

In the context of assessing the convergence of the MLMC algorithm in a standardized manner, the authors in [4] theoretically prove, with specific assumptions on the deterministic solution, two error bounds for the multilevel sample mean and variance that may be directly compared. But no practical normalized error estimates are defined for interpretation of the complexities between the moments. On the other hand, relative error estimates of the MC and MLMC methods, where the total MSE of a given moment is normalized by the square of its own statistic, are considered in [35]. However, such relative errors are not fully scale-invariant. For example, in the MC estimation of the mean, any additive linear transformation of the QoI changes the ratio of absolute MSE to the squared mean in a dimension-dependent manner. Interestingly, in the MC variance estimation, one does obtain a fully dimensionless relative error under linear scaling and translation, when normalized by the square of the variance. However, the resulting error estimate becomes highly sensitive to the tail behaviour (kurtosis) of the QoI distribution. To tackle this issue, in this article, we propose novel normalized mean square error (NMSE) estimates for both the MC mean and variance estimations, based on which the MLMC counterparts are derived. The newly introduced relative errors are statistically defined using h-statistics, with chosen normalizing factors that are finite and unbiased. They ensure that the total MSEs of the MC and the MLMC algorithms are fully scale-invariant under any linear transformation (scaling and addition) of QoI, and remain robust to variations in distributional characteristics. Therefore, the proposed NMSEs enable easier interpretation of statistical accuracy and efficiency between the MC and MLMC algorithms for the estimation of both mean and variance, and across different scales.

The other objective of this paper is to investigate the applicability of the scale-invariant MC and MLMC method to linear elliptic problems described by stochastic material parameters representing both heterogeneity and uncertain symmetries. As an example, we consider the linear-elastic material model of the human femoral bone, whose constitutive law is assumed to be uncertain. In particular, bone tissue is not only a highly heterogeneous material but also anisotropic, where the material symmetry is described as uncertain due to the lack of conclusive verification of the class of elastic symmetry it belongs to [31,53]. Hence, the entire elasticity tensor is constructed as random with a predefined elastic symmetry in the mean (orthotropy [18] in this case) and triclinic symmetries in each stochastic realization. Consequently, the positive-definite random elasticity matrices are modelled as matrix-valued random fields, as proposed in [61,62]. In this work, we restrict ourselves to material uncertainties only, assuming that the remaining model parametrization is deterministic. However, the numerical approach is general enough to be employed for other types of uncertainties as well.

The paper is organised as follows: In Section 2, we describe the problem, whereas the theoretical procedures of scale-invariant MC mean and variance estimates are elaborated in Section 3. Following this, in Section 4, the normalized versions of the MLMC mean and variance estimators are detailed. A deterministic and stochastic setting of the linear elastic material model, along with the stochastic material modelling, is given in Section 5. In Section 6, we visualize the considered stochastic material model when implemented on a two-dimensional proximal femur, and detail the accuracy and efficiency of the normalized MC and MLMC. Finally, the conclusions are drawn in Section 7.

## 2. Problem description

Let us consider a physical system occupying the spatial domain $\mathcal{G} \subset \mathbb{R}^d$ in a $d$-dimensional Euclidean space, modelled by an abstract equilibrium equation:

$$\mathcal{A}(q(x), u(x)) = f(x). \tag{2}$$

Here, $u(x) \in \mathcal{U}$ describes the state of the system at a spatial point $x \in \mathcal{G}$ lying in a Hilbert space $\mathcal{U}$ (for the sake of simplicity), $\mathcal{A}$ is a (possibly non-linear) operator modelling the physics of the system, and $f \in \mathcal{U}^*$, the dual space of $\mathcal{U}$, is some external influence (action/excitation/loading). Furthermore, we assume that the model depends on the parameter set $q \in \mathcal{Q}$ and that it is accompanied by appropriate boundary and/or initial conditions. Note that, for brevity, the model in the previous equation describes the physical system only in a spatial domain $x$, whereas one may also generalise this to time-dependent processes.

The uncertainty in the previous equation may arise due to the randomness in external influence $f$, initial or boundary conditions, geometry $\mathcal{G}$, as well as the coefficient of operator $\mathcal{A}$, i.e., parameter $q$. Although the theory presented further does not depend on this choice and is general enough to cover all of the mentioned (single or combination of) cases, this article focuses on incorporating stochasticity only in coefficient $q$. In the theory of continuum solid mechanics, the parameter $q$ represents one of the very well-known physical phenomena, such as elasticity [49], which is detailed as an example in Section 5. Here, we assume that $q$ is modelled as a random field $q(x, \omega)$ with finite second-order moments on a probability space $(\Omega, \mathfrak{F}, \mathbb{P})$. Following this, Eq. (2) rewrites to a stochastic form:

$$\mathcal{A}(q(x, \omega), u(x, \omega)) = f(x), \tag{3}$$

which further is to be solved for $u(x, \omega) \in \mathcal{U} \otimes L_2(\Omega, \mathfrak{F}, \mathbb{P})$.

Often, when the response of a model is uncertain, one is interested in obtaining its relevant information via its corresponding statistics. Analytically, the statistical moments of the solution $u \equiv u(x, \omega)$ can be represented, by rewriting Eq. (2), as $p$–th central moment:

$$\mu_p(u) = \mathbb{E}\left((u - \mathbb{E}(u))^p\right) = \int_\Omega (u - \mathbb{E}(u))^p \mathbb{P}(\mathrm{d}\omega). \tag{4}$$

The objective of this study is to determine the mean and variance only after an appropriate deterministic discretization of the problem in Eq. (3) is presented.

Due to spatial and stochastic dependence, the solution $u \equiv u(x,\omega)$ in Eq. (3) is first discretized in a spatial domain, i.e., we search for the solution in a finite subspace $\mathcal{U}_h \subset \mathcal{U}$. After rewriting the problem in Eq. (3) in a variational form, the spatial discretization $u_h(x,\omega) \in \mathcal{U}_h$—$h$ being the discretization parameter—can take the finite element form on a sufficiently fine spatial mesh $\mathcal{T}_h$ [1] [70]. The expectation functional in Eq. (4) then rewrites to:

$$\mu_p(u_h) = \mathbb{E}((u_h - \mathbb{E}(u_h))^p) = \int_\Omega (u_h - \mathbb{E}(u_h))^p \mathbb{P}(\mathrm{d}\omega), \tag{5}$$

in which the exact solution $u$ is substituted by a semi-discretized solution $u_h \equiv u_h(x,\omega)$.

## 3. Monte Carlo method

Due to the complexity of integrating Eq. (5) analytically, in this paper, the focus is on the traditional sampling-based Monte Carlo (MC) method [16,22,27,47].

### 3.1. Monte Carlo estimation of mean

For an unbiased estimate of the statistic $g(\omega)$, one may take a symmetric function

$$\widehat{\mu}^{\mathrm{MC}}(g) = \frac{1}{N}(g(\omega_1) + g(\omega_2) + \cdots + g(\omega_N)), \tag{6}$$

meaning that the estimate does not depend on the order in which observations were taken. The existence and uniqueness of such a choice are given in [25]. Under the assumption that each sample $u_h(x,\omega_i)$ comes from the identical distribution as $u_h(x,\omega)$ and by use of Eq. (6), one may reformulate the MC estimate of the mean $\mu(u_h)$ as

$$\mathbb{E}(u_h) \approx \widehat{\mu}^{\mathrm{MC}}(u_h) = \frac{1}{N}\sum_{i=1}^{N} u_h(x,\omega_i), \tag{7}$$

where $N > 1$ is the sample size of the random field $u_h(x,\omega)$ at a spatial location $x$. Following this, the approximation error of the MC-based mean estimate $\widehat{\mu}^{\mathrm{MC}} \equiv \widehat{\mu}^{\mathrm{MC}}(u_h)$ compared to the exact mean reads

$$e(\widehat{\mu}^{\mathrm{MC}}) = \widehat{\mu}^{\mathrm{MC}} - \mathbb{E}(u), \tag{8}$$

which further can be rewritten as

$$e(\widehat{\mu}^{\mathrm{MC}}) = \widehat{\mu}^{\mathrm{MC}} - \mathbb{E}(u_h) + \mathbb{E}(u_h) - \mathbb{E}(u).$$

Thus, the mean square error (MSE) reads

$$\begin{aligned}\mathrm{MSE}(\widehat{\mu}^{\mathrm{MC}}) &:= \mathbb{E}(e^2) = \mathbb{E}((\widehat{\mu}^{\mathrm{MC}} - \mathbb{E}(u_h) + \mathbb{E}(u_h) - \mathbb{E}(u))^2)\\ &= \mathbb{E}((\widehat{\mu}^{\mathrm{MC}} - \mathbb{E}(u_h))^2) + (\mathbb{E}(u_h) - \mathbb{E}(u))^2\\ &\quad + 2(\mathbb{E}(u_h) - \mathbb{E}(u))(\mathbb{E}(\widehat{\mu}^{\mathrm{MC}}) - \mathbb{E}(u_h)),\end{aligned}$$

in which $\widehat{\mu}^{\mathrm{MC}}$ is assumed to be an unbiased estimator; meaning that $\mathbb{E}(\widehat{\mu}^{\mathrm{MC}}) = \mathbb{E}(u_h)$, and also, $\mathbb{E}((\widehat{\mu}^{\mathrm{MC}} - \mathbb{E}(u_h))^2) = \mathbb{V}\mathrm{ar}(\widehat{\mu}^{\mathrm{MC}}) = \mathbb{V}\mathrm{ar}(u_h)/N$, via the central limit theorem [15]. Therefore, the previous equation reduces to:

$$\mathrm{MSE}(\widehat{\mu}^{\mathrm{MC}}) = \frac{\mathbb{V}\mathrm{ar}(u_h)}{N} + (\mathbb{E}(u_h) - \mathbb{E}(u))^2. \tag{9}$$

The first term in the previous equation is the sampling error, which varies inversely to the sample size $N$. On the other hand, the second term describes the square of the spatial discretization error, whose change is proportional to the element size $h$.

[1] Other types of discretization techniques can be considered as well

**Scale-invariant error estimator for MC mean estimation**: The MSE in Eq. (9) is an absolute error estimate and, hence, scale-dependent. For example, if each observation $u_h(x,\omega_i)$ is linearly transformed in the form $a\,u_h(x,\omega_i)+b$, in which $a, b \in \mathbb{R}$ are constants, then the Monte Carlo error in Eq. (9) will be affected in a square proportional manner—as

$$\mathbb{V}\mathrm{ar}(a\,u_h + b) = a^2\,\mathbb{V}\mathrm{ar}(u_h) \tag{10}$$

and

$$(\mathbb{E}(a\,u_h + b) - \mathbb{E}(a\,u + b))^2 = a^2(\mathbb{E}(u_h) - \mathbb{E}(u))^2. \tag{11}$$

This leads to an interpretability issue, which is not suitable in practical applications—for example, when transforming the units of temperature from Kelvin to Fahrenheit or displacement field from millimetres to metres. Moreover, an additional complication arises when comparing the convergence behaviour of different MC moments under a fixed MSE. Specifically, under a linear transformation of the form defined in the previous equations, the MSE of the MC mean estimator scales with $a^2$, while the MSE of the MC variance estimator scales with $a^4$—more information on this to follow in Section 3.2. To address this limitation, it is necessary to define a scale-invariant version of MSE, defined in Eq. (9), by incorporating a statistical normalization factor $\lambda_m$, which must satisfy the following properties:

(a) It must be greater than zero and finite i.e., $0 < \lambda_m < \infty$;
(b) It should satisfy the scaling condition $\lambda_m(a\,u_h + b) = a^2\,\lambda_m$, thereby achieving a completely dimensionless MSE;
(c) It must be chosen such that the sampling error of the scale-invariant error estimator remains robust and fully invariant with respect to variations in the properties of the distribution of solution $u_h$.

**$\mu^2$ as a normalizer**: It is important to highlight that the common practice of using the squared mean value $\mu^2 = \mathbb{E}(u_h)^2$ as the standardizing quantity $\lambda_m$ satisfies the second criterion listed above—namely, achieving a dimensionless MSE—only in the case of multiplicative scale-change of the form $\mu(au_h)^2 = a^2\mu(u_h)^2$. However, this choice fails under additive transformations, as $\mu(u_h + b)^2 = (\mu(u_h) + b)^2$; thereby, violating the required invariance. Consequently, $\mu^2$ cannot be considered a suitable normalizing factor for constructing a truly scale-invariant error metric.

We propose using the variance of the solution, $\lambda_m = \mathbb{V}\mathrm{ar}(u_h)$, as a normalizer. Provided that $0 < \mathbb{V}\mathrm{ar}(u_h) < \infty$, the factor $\lambda_m$ satisfies the second condition outlined above (see Eq. (10)). With this, the new standardized error estimate is defined as

$$\hat{e}(\widehat{\mu}^{\mathrm{MC}}) := \frac{e(\widehat{\mu}^{\mathrm{MC}})}{\sqrt{\mathbb{V}\mathrm{ar}(u_h)}}. \tag{12}$$

Finally, by normalizing Eq. (9), one obtains the squared error estimate

$$\mathrm{NMSE}(\widehat{\mu}^{\mathrm{MC}}) := \mathbb{E}(\hat{e}^2) = \frac{1}{N} + \frac{(\mathbb{E}(u_h) - \mathbb{E}(u))^2}{\mathbb{V}\mathrm{ar}(u_h)}, \tag{13}$$

in which the first term is the scale-invariant sampling accuracy, and the second term represents the new normalized squared discretization error. In the previous equation, the resulting NMSE remains fully dimensionless under both multiplicative and additive scale transformations. Furthermore, the normalized sampling error is only proportional to $1/N$—making it invariant to the variations in variance of solution $u_h$. In other words, for a given total NMSE and a fixed spatial resolution $h$, the computational cost of the MC mean estimator $\widehat{\mu}^{\mathrm{MC}}$ (see Proposition. 3.1) is solely dependent on the number of MC samples $N$, thereby fulfilling the third listed requirement.

The computational complexity of the scale-invariant MC mean estimate is similar to the conventional procedure, as shown in [9], but it must be redefined with respect to the total NMSE, which is given below.

**Proposition 3.1.** *Let us consider $c_\alpha, c_\gamma, \alpha$ and $\gamma$ as positive constants, and then one may define the error bounds as follows, where*

*(i) The deterministic error decays as* $\dfrac{|\mathbb{E}(u_h) - \mathbb{E}(u)|}{\sqrt{\lambda_m}} \leq c_\alpha h^\alpha$,

*(ii) The computational cost to determine a single realization of* $u_h(x,\omega)$ *is given by* $C(u_h) \leq c_\gamma h^{-\gamma}$.

*Then, for any* $0 < \mathring{\varepsilon} < e^{-1}$, *the Monte Carlo (MC) mean estimator* $\widehat{\mu}^{MC}$ *with* $N = \mathcal{O}(\mathring{\varepsilon}^{-2})$ *and* $h = \mathcal{O}(\mathring{\varepsilon}^{1/\alpha})$ *satisfies the normalized mean square error* $NMSE(\widehat{\mu}^{MC}) < \mathring{\varepsilon}^2$. *Therefore, the corresponding computational cost of MC mean estimation is*

$$C(\widehat{\mu}^{MC}) \leq c\, \mathring{\varepsilon}^{-2-\gamma/\alpha},$$

*where* $c > 0$ *is a positive constant.*

**Unbiased estimation of $\mathbb{V}$ar**: In general, the true variance $\mathbb{V}\mathrm{ar}(u_h)$ is unknown, and must be estimated. For this purpose, one may use the estimator $\widehat{\mathbb{V}}\mathrm{ar}^{\mathrm{MC}}(u_h)$, which is determined by the MC procedure with $N$ random draws. However, in such a case, the use of a symmetric function as in Eq. (6) does not lead to an unbiased estimator. To address this limitation, we make use of h-statistics for estimating the central moments $\mu_p$, which are not only unbiased but also symmetric, and possess minimal variance [14,51]. Further details on h-statistics for univariate central moments, particularly the second-order moment (variance), are provided in Appendix A. Based on this, we rewrite Eq. (12) by estimating $\lambda_m = \mathbb{V}\mathrm{ar}(u_h)$ using the second h-statistic, denoted as $\widehat{\lambda}_m = \widehat{\mathrm{h}}_2^{\mathrm{MC}}$ (see Eq. (72)), resulting:

$$\widehat{\mathring{\varepsilon}}(\widehat{\mu}^{\mathrm{MC}}) = \frac{e(\widehat{\mu}^{\mathrm{MC}})}{\sqrt{\widehat{\mathrm{h}}_2^{\mathrm{MC}}}}. \tag{14}$$

Finally, one obtains a new formulation of the estimated scale-invariant MSE (from Eq. (13)) in the form:

$$\mathrm{NMSE}(\widehat{\mu}^{\mathrm{MC}}) = \frac{1}{N} + \frac{(\mathbb{E}(u_h) - \mathbb{E}(u))^2}{\widehat{\mathrm{h}}_2^{\mathrm{MC}}}. \tag{15}$$

### 3.2. Monte Carlo estimation of variance

The previously derived estimate of variance by h-statistics in Eq. (72) is characterized by an approximation error given as

$$e(\widehat{\mathrm{h}}_2^{\mathrm{MC}}) = \widehat{\mathrm{h}}_2^{\mathrm{MC}} - \mathbb{V}\mathrm{ar}(u_h) + \mathbb{V}\mathrm{ar}(u_h) - \mathbb{V}\mathrm{ar}(u),$$

in which $\mathbb{V}\mathrm{ar}(u)$ denotes the exact variance of the solution $u$. Following this, the MSE of the variance estimator reads:

$$\begin{aligned}\mathrm{MSE}(\widehat{\mathrm{h}}_2^{\mathrm{MC}}) &:= \mathbb{E}(e^2) = \mathbb{E}((\widehat{\mathrm{h}}_2^{\mathrm{MC}} - \mathbb{V}\mathrm{ar}(u_h) + \mathbb{V}\mathrm{ar}(u_h) - \mathbb{V}\mathrm{ar}(u))^2)\\ &= \mathbb{E}((\widehat{\mathrm{h}}_2^{\mathrm{MC}} - \mathbb{V}\mathrm{ar}(u_h))^2) + (\mathbb{V}\mathrm{ar}(u_h) - \mathbb{V}\mathrm{ar}(u))^2\\ &\quad + 2(\mathbb{V}\mathrm{ar}(u_h) - \mathbb{V}\mathrm{ar}(u))(\mathbb{E}(\widehat{\mathrm{h}}_2^{\mathrm{MC}}) - \mathbb{V}\mathrm{ar}(u_h)).\end{aligned}$$

Having that $\widehat{\mathrm{h}}_2^{\mathrm{MC}}$ is an unbiased estimator, meaning that $\mathbb{E}(\widehat{\mathrm{h}}_2^{\mathrm{MC}}) = \mathbb{V}\mathrm{ar}(u_h)$, one may further rewrite the previous equation to

$$\mathrm{MSE}(\widehat{\mathrm{h}}_2^{\mathrm{MC}}) = \mathbb{V}\mathrm{ar}(\widehat{\mathrm{h}}_2^{\mathrm{MC}}) + (\mathbb{V}\mathrm{ar}(u_h) - \mathbb{V}\mathrm{ar}(u))^2. \tag{16}$$

In the previous equation, the variance of the second-order statistic $\widehat{\mathrm{h}}_2^{\mathrm{MC}}$, which also represents the normalizing constant $\widehat{\lambda}_m$, is derived by [8] and reads

$$\mathbb{V}\mathrm{ar}(\widehat{\mathrm{h}}_2^{\mathrm{MC}}) := \frac{1}{N}\left(\mu_4(u_h) - \frac{\mu_2(u_h)^2(N-3)}{N-1}\right), \tag{17}$$

where $\mu_2(u_h)$ and $\mu_4(u_h)$ represent the second and fourth central population moments, respectively. For brevity, one may further rewrite the previous equation as

$$\mathbb{V}\mathrm{ar}(\widehat{\mathrm{h}}_2^{\mathrm{MC}}) := \frac{\mathbb{V}_2(u_h)}{N}, \tag{18}$$

where

$$\mathbb{V}_2(u_h) := \mu_4(u_h) - \frac{\mu_2(u_h)^2(N-3)}{N-1}. \tag{19}$$

Consequently, by substituting Eq. (18) in Eq. (16), one obtains

$$\mathrm{MSE}(\widehat{\mathrm{h}}_2^{\mathrm{MC}}) = \frac{\mathbb{V}_2(u_h)}{N} + (\mathbb{V}\mathrm{ar}(u_h) - \mathbb{V}\mathrm{ar}(u))^2. \tag{20}$$

Here, the first term defines the statistical error, which is inversely proportional to the number of samples $N$, and the second part represents the square of the discretization error, directly proportional to parameter $h$.

**Scale-invariant error estimator for MC variance estimation**: Similar to MSE formulation for the MC mean estimator in Eq. (9), the error estimate for the MC variance estimator, as presented in Eq. (20), is inherently scale-dependent. That is, under a linear transformation of the form $au_h + b$, as discussed in Section 3.1, the corresponding MSE exhibits a bi-quadratic scaling behaviour. This is due to the characteristics of the central moments involved in Eq. (19):

$$\mu_4(a\,u_h + b) = a^4\,\mu_4(u_h) \tag{21}$$

and

$$\mu_2{}^2(a\,u_h + b) = a^4\,\mu_2{}^2(u_h). \tag{22}$$

Therefore, it is clear that

$$\mathbb{V}_2(a\,u_h + b) = a^4\,\mathbb{V}_2(u_h) \tag{23}$$

and also that

$$(\mathbb{V}\mathrm{ar}(a\,u_h + b) - \mathbb{V}\mathrm{ar}(a\,u + b))^2 = a^4(\mathbb{V}\mathrm{ar}(u_h) - \mathbb{V}\mathrm{ar}(u))^2. \tag{24}$$

Therefore, similar to the standardizing entity $\lambda_m$ in Section 3.1, the normalization entity for the MC variance estimation, denoted by $\lambda_v$, must satisfy the following criteria:

(a) It must be finite and greater than zero, i.e., $0 < \lambda_v < \infty$;
(b) It should satisfy the scaling condition $\lambda_v(a\,u_h + b) = a^4\,\lambda_v$—to achieve a completely dimensionless MSE;
(c) It must be chosen such that the sampling error of the scale-invariant error estimator remains robust and fully invariant with respect to variations in the properties of the distribution of solution $u_h$.

**$\mathbb{V}$ar$^2$ $\equiv$ $\mu_2{}^2$ as a normalizer**: If one considers $\lambda_v = \mu_2{}^2$ as a standardizing factor, the MSE in Eq. (20) transforms to

$$\mathrm{NMSE}(\widehat{\mathrm{h}}_2^{\mathrm{MC}}) = \frac{\mathbb{V}_2(u_h)}{\mu_2{}^2\,N} + \frac{(\mathbb{V}\mathrm{ar}(u_h) - \mathbb{V}\mathrm{ar}(u))^2}{\mu_2{}^2}. \tag{25}$$

Provided that $0 < \lambda_v < \infty$, the normalization by $\mu_2^2$ enables the NMSE to be invariant under any linear transformation (as established in Eq. (22)). This satisfies the second requirement in the above list. However, regarding the third criterion—robustness of the normalized

sampling error to changes in the distribution of $u_h$—further analysis is required. The normalized sampling error, denoted by $\mathring{\varepsilon}_s$, derived from Eq. (19), is given as:

$$\mathring{\varepsilon}_s = \frac{1}{N}\left(\frac{\mu_4}{{\mu_2}^2} - \frac{N-3}{N-1}\right), \tag{26}$$

where we have suppressed the $u_h$ notation for clarity. In the asymptotic regime where $N \gg 1$, the ratio $(N-3)/(N-1) \to 1$, and the expression simplifies to:

$$\mathring{\varepsilon}_s = \frac{1}{N}\left(\frac{\mu_4}{{\mu_2}^2} - 1\right), \tag{27}$$

in which the term $\mu_4/\mu_2^2$ represents the standardized fourth central moment, kurtosis, denoted by $\text{Kurt} = \mu_4/{\mu_2}^2$ [13]. The normalized sampling error $\mathring{\varepsilon}_s$ depends not only on the number of samples $N$ but also on the kurtosis of the solution distribution. This dependence becomes particularly significant for heavy-tailed distributions, which exhibit high kurtosis values.

For example, in the case of a (symmetric) Gaussian distribution, the kurtosis is $\text{Kurt} = 3$ [13]. In contrast, for a log-normal distribution, the kurtosis is given by $\text{Kurt} = \rho^4 + 2\rho^3 + 3\rho^2 - 3$ [12], where $\rho = \exp(\mathbb{V}\text{ar})$ and $\mathbb{V}\text{ar}$ is the variance of the underlying Gaussian distribution. If $\mathbb{V}\text{ar} = 1$, then $\text{Kurt} \approx 114$. This implies that, for a fixed total NMSE and spatial resolution $h$, the computational effort required to estimate the MC variance $\widehat{\text{h}}_2^{\text{MC}}$ is twice that of the mean estimator $\widehat{\mu}^{\text{MC}}$ for a Gaussian-distributed $u_h$ (since $\mathring{\varepsilon}_s = 2/N$). However, for a log-normal distribution, the cost increases dramatically—by a factor of approximately 113 (i.e., $\mathring{\varepsilon}_s \approx 113/N$)—relative to the mean estimator. This violates the third criteria in the list, and therefore, undermines the effectiveness of $\mathbb{V}\text{ar}^2 \equiv {\mu_2}^2$ as a normalizer in practical applications.

Following this, we consider $\lambda_v = \mathbb{V}_2(u_h)$ as a normalizer for MC variance estimation. Given that $0 < \mathbb{V}_2(u_h) < \infty$, and based on the scaling relation in Eq. (23), it is evident that $\mathbb{V}_2(u_h)$ satisfies the second outlined requirement. By normalizing the MSE with $\mathbb{V}_2(u_h)$ in Eq. (20), we obtain the scale-invariant form of the MSE for the MC variance estimator:

$$\text{NMSE}(\widehat{\text{h}}_2^{\text{MC}}) := \mathbb{E}(\mathring{e}^2) = \frac{1}{N} + \frac{(\mathbb{V}\text{ar}(u_h) - \mathbb{V}\text{ar}(u))^2}{\mathbb{V}_2(u_h)}. \tag{28}$$

Here, the first term represents the scale-invariant sampling error, which decays proportionally to $1/N$ and is independent of the statistical characteristics of the solution $u_h$, thereby fulfilling the third requirement. The second term denotes the normalized squared discretization error.

Furthermore, based on the computational complexity analysis for MC variance estimation discussed in [35], we henceforth recast the convergence analysis in terms of the normalized MSE formulation.

**Proposition 3.2.** *Let us consider $c_\alpha, c_\gamma, \alpha$ and $\gamma$ as positive constants, and then one may define the error bounds, such that*

*(i) The deterministic error is bounded by* $\dfrac{|\mathbb{V}ar(u_h) - \mathbb{V}ar(u)|}{\sqrt{\lambda_v}} \leq c_\alpha h^\alpha$,

*(ii) The cost to compute each sample of $u_h(x,\omega)$ is given as* $C(u_h) \leq c_\gamma h^{-\gamma}$.

*Then, for any $0 < \mathring{\epsilon} < e^{-1}$, the Monte Carlo (MC) variance estimator $\widehat{h}_2^{MC}$ with $N = \mathcal{O}(\mathring{\epsilon}^{-2})$ and $h = \mathcal{O}(\mathring{\epsilon}^{1/\alpha})$ satisfies the normalized mean square error $NMSE(\widehat{h}_2^{MC}) < \mathring{\epsilon}$. Therefore, the corresponding computational cost of MC variance estimation is given as*

$$C(\widehat{h}_2^{MC}) \leq c\, \mathring{\epsilon}^{-2-\gamma/\alpha},$$

*where $c > 0$ is a positive constant.*

**Unbiased estimation of $\mathbb{V}_2$:** The quantity $\mathbb{V}_2(u_h)$ in Eq. (28) is analytical and not known. To determine the unbiased estimate of $\mathbb{V}_2(u_h)$, one must obtain unbiased estimates of entities $\mu_4(u_h)$ and $\mu_2(u_h)^2$, which are detailed in Appendix B. Therefore, by substituting $\mu_4 \approx \text{h}_4^{\text{MC}}$ and $\mu_2^2 \approx \text{h}_{\{2,2\}}^{\text{MC}}$ from Eqs. (74) and (75), respectively, in Eq. (19), one obtains the unbiased estimate: $\widehat{\lambda}_v := \widehat{\mathbb{V}}_2^{\text{MC}}(u_h)$—of $\mathbb{V}_2(u_h)$. A further analysis of the stochastic convergence of the estimator $\widehat{\mathbb{V}}_2^{\text{MC}}(u_h)$ is provided in Appendix C. Finally, the MSE in Eq. (28) is re-described as,

$$\text{NMSE}(\widehat{\text{h}}_2^{\text{MC}}) := \mathbb{E}(\mathring{e}^2) = \frac{1}{N} + \frac{(\mathbb{V}\text{ar}(u_h) - \mathbb{V}\text{ar}(u))^2}{\widehat{\mathbb{V}}_2^{\text{MC}}(u_h)}. \tag{29}$$

## 4. Multilevel Monte Carlo method

The MSEs of MC mean and variance estimators in Eqs. (9) and (20), respectively, signify that to attain an overall higher level of total accuracy, one requires a very fine resolution of the finite element mesh and a very large number of MC samples. This demands a tremendous amount of computational effort, making the algorithm practically infeasible. Therefore, the desired moments are further estimated by a variance reduction technique in a multilevel fashion following [20,21,26].

### 4.1. Multilevel Monte Carlo estimation of mean

Let $\{l = 0, 1, 2 \ldots, L\}$ be a generalised increasing sequence—in the context of decreasing element size $h$—of nested meshes $\mathcal{P}_l$, a regular (non-degenerate) partition of the computational domain $\mathcal{G}$ of the problem described in Eq. (2). Here, $l$ denotes the mesh level, and $L$ represents the finest mesh. The goal of the MLMC method is to determine the statistics (such as the mean in this case) of the solution $u_{h_L}(\boldsymbol{x}, \omega)$ on the finest level $L$. To this end, by exploiting the linearity of the expectation operator, one may express the MLMC (for brevity, denoted by ML) mean estimate of the mean $\mu(u_{h_L})$ using a set of samples $\{N_l\} := \{N_0, N_1, \ldots, N_L\}$ as [21]

$$\begin{aligned}\widehat{\mu}^{\text{ML}}(u_{h_L,\{N_l\}}) &:= \widehat{\mu}^{\text{MC}}(u_{h_0,N_0}) + \sum_{l=1}^{L} \widehat{\mu}^{\text{MC}}(u_{h_l,N_l} - u_{h_{l-1},N_l}) \\ &= \sum_{l=0}^{L} \widehat{\mu}^{\text{MC}}(Y_l).\end{aligned} \tag{30}$$

Here, $\widehat{\mu}^{\text{MC}}(u_{h_0,N_0})$ is the MC estimator of mean $\mu(u_{h_0})$ on level $l = 0$ using $N_0$ samples, and $\widehat{\mu}^{\text{MC}}(u_{h_l,N_l} - u_{h_{l-1},N_l})$ represents the approximation of mean $\mu(u_{h_l} - u_{h_{l-1}})$ with $l > 0$ and $N_l$ samples. Furthermore, for $l = 0$, $Y_0 = u_{h_0,N_0}$; else, $Y_l := u_{h_l,N_l} - u_{h_{l-1},N_l}$. Note that the individual term $Y_l$, $l \geq 0$, is sampled independently, and when $l > 0$, the quantities $u_{h_l,N_l}$ and $u_{h_{l-1},N_l}$ in $Y_l$ are considered to be strongly correlated—meaning that $u_{h_l,N_l}$ and $u_{h_{l-1},N_l}$ are sampled from the same random seed.

As the mean estimate on the finest level $\widehat{\mu}^{\text{ML}} \equiv \widehat{\mu}^{\text{ML}}(u_{h_L,\{N_l\}})$ is obtained as the telescopic sum of the differences of MC mean estimates on the coarser levels, the MSE of $\widehat{\mu}^{\text{ML}}$, corresponding to Eq. (9), takes the form (see [9]):

$$\text{MSE}(\widehat{\mu}^{\text{ML}}) = \sum_{l=0}^{L} \frac{\mathbb{V}\text{ar}(Y_l)}{N_l} + (\mathbb{E}(u_{h_L}) - \mathbb{E}(u))^2. \tag{31}$$

The above error consists of two terms: the variance of the estimator $\widehat{\mu}^{\text{ML}}$ on the left and the square of the spatial discretization error on the right.

**Scale-invariant error estimator for MLMC mean estimation:** Similar to the MC error estimate for the mean in Eq. (9), the MSE of the MLMC mean estimator in Eq. (31) represents an absolute error estimate. As a result, the convergence of the MLMC algorithm also strongly depends on the solution magnitude. To address this, based on the normalizer $\lambda_m$ defined in Section 3.1, and particularly the NMSE in Eq. (13), we consider $\lambda_m = \mathbb{V}\text{ar}(u_{h_L})$ as a standardizing entity, suggesting a new scale-invariant MSE estimate:

$$\text{NMSE}(\widehat{\mu}^{\text{ML}}) = \frac{1}{\mathbb{V}\text{ar}(u_{h_L})}\left(\sum_{l=0}^{L} \frac{\mathbb{V}\text{ar}(Y_l)}{N_l}\right) + \frac{(\mathbb{E}(u_{h_L}) - \mathbb{E}(u))^2}{\mathbb{V}\text{ar}(u_{h_L})}. \tag{32}$$

However, the normalization term $\mathbb{V}\mathrm{ar}(u_{h_L})$, which is the variance of the solution $u_h$ on finest mesh level $L$, is generally not known. Hence, under the assumption that for any value of $l$, $\mathbb{V}\mathrm{ar}(u_{h_l})$ will be approximately constant [9,21], one may substitute $\mathbb{V}\mathrm{ar}(u_{h_L})$ with $\mathbb{V}\mathrm{ar}(u_{h_0})$, which is based on the coarsest mesh at $l = 0$. Furthermore, under the consideration that $\widehat{\lambda}_m := \widehat{\mathrm{h}}_2^{\mathrm{MC}}(u_{h_0,N_0})$ is the MC estimator of $\mathbb{V}\mathrm{ar}(u_{h_0})$, and by defining the MC variance estimate of $\mathbb{V}\mathrm{ar}(Y_l) := \mathbb{V}\mathrm{ar}(u_{h_l} - u_{h_{l-1}})$ via $\widehat{\mathrm{h}}_2^{\mathrm{MC}}(Y_l)$, Eq. (32) therefore transforms to

$$\mathrm{NMSE}(\widehat{\mu}^{\mathrm{ML}}) = \frac{1}{\widehat{\mathrm{h}}_2^{\mathrm{MC}}(u_{h_0,N_0})}\left(\sum_{l=0}^{L}\frac{\widehat{\mathrm{h}}_2^{\mathrm{MC}}(Y_l)}{N_l}\right) + \frac{(\mathbb{E}(u_{h_L}) - \mathbb{E}(u))^2}{\widehat{\mathrm{h}}_2^{\mathrm{MC}}(u_{h_0,N_0})}. \tag{33}$$

The first entity here represents the estimated scale-invariant multi-level sampling error, whereas the second term defines the dimensionless squared discretization error. Therefore, to attain an overall normalized mean square error $\hat{\epsilon}^2$, it is sufficient that both terms are less than $\hat{\epsilon}^2/2$. Note that the equal splitting of error $\hat{\epsilon}^2$ is not a requirement and can also be set otherwise—for more information on this, see [10,24].

If $C_l \equiv C(Y_l)$ is the computational cost of determining a single MC sample of $Y_l$, then the overall cost of the estimator $\widehat{\mu}^{\mathrm{ML}}$ is given as $C(\widehat{\mu}^{\mathrm{ML}}) := \sum_{l=0}^{L} N_l C_l$. Here, the optimum number of samples $N_l$ on each level $l$ is evaluated by solving an optimisation problem such that the normalized sampling error is less than $\hat{\epsilon}^2/2$. As a result, the cost function

$$f(N_l) = \arg\min_{N_l} \sum_{l=0}^{L}\left(N_l C_l + \tau\frac{\widehat{\mathrm{h}}_2^{\mathrm{MC}}(Y_l)}{\widehat{\lambda}_m N_l}\right) \tag{34}$$

is minimised, due to which the optimal samples $N_l$ are calculated as

$$N_l = \tau\left(\frac{\widehat{\mathrm{h}}_2^{\mathrm{MC}}(Y_l)}{\widehat{\lambda}_m C_l}\right)^{\frac{1}{2}}. \tag{35}$$

Here, $\tau$ is the Lagrange multiplier, determined by

$$\tau = \frac{2}{\hat{\epsilon}^2}\sum_{l=0}^{L}\left(\frac{\widehat{\mathrm{h}}_2^{\mathrm{MC}}(Y_l)C_l}{\widehat{\lambda}_m}\right)^{\frac{1}{2}}. \tag{36}$$

On the other hand, the set of mesh levels $\{l = 0, 1, 2 \ldots, L\}$ maybe optimally chosen (in a geometric or non-geometric sequence) for a given deterministic error. Typically, for PDE-based applications, the choice of coarse mesh $l = 0$ depends on the regularity of the solution $u(x)$ [9]. Following this, the finer mesh selection is based on an apriori mesh convergence study, where the finest mesh level $l = L$ can be fixed or can be adaptively selected [21] during the MLMC mean computation. For a more detailed discussion on the optimal selection of mesh hierarchies, we refer the readers to [24].

The arguments for determining the total computational cost of the normalized MLMC mean estimator follow the classical MLMC procedure [20,26]. However, the difference is that the cost is described with respect to NMSE instead of MSE.

**Proposition 4.1.** *Let us consider the positive constants $c_\alpha, c_\beta, c_\gamma, \alpha, \beta, \gamma$, given that $\alpha \geq \frac{1}{2}\min(\beta, \gamma)$. Then, we consider the following error bounds, where*

$$\begin{aligned}
&(i)\ \textit{The deterministic error decays as } \frac{|\mathbb{E}(u_{h_l}) - \mathbb{E}(u)|}{\sqrt{\lambda_m}} \leq c_\alpha h_l^\alpha,\\
&(ii)\ \textit{The decay of variance is bounded by } \frac{\widehat{h}_2^{MC}(Y_l)}{\lambda_m} \leq c_\beta h_l^\beta,\\
&(iii)\ \textit{The computational cost to determine a single realization of } Y_l\\
&\quad\textit{is given as } C(Y_l) \leq c_\gamma h_l^{-\gamma}.
\end{aligned} \tag{37}$$

*Then, there exists another positive constant $c$, such that for any $0 < \hat{\epsilon} < e^{-1}$, the multilevel Monte Carlo (MLMC) mean estimator satisfies the normalized mean square error $\mathrm{NMSE}(\widehat{\mu}^{ML}) < \hat{\epsilon}^2$. Finally, the total computational cost of MLMC mean estimation is bounded by*

$$C(\widehat{\mu}^{ML}) \leq \begin{cases} c\,\hat{\epsilon}^{-2}, & \beta > \gamma,\\ c\,\hat{\epsilon}^{-2}(\log\hat{\epsilon})^2, & \beta = \gamma,\\ c\,\hat{\epsilon}^{-2-(\gamma-\beta)/\alpha}, & \beta < \gamma. \end{cases} \tag{38}$$

Based on the values of $\beta$ and $\gamma$, one may also understand the major cost contributor amongst the sequence of mesh levels. If $\beta > \gamma$, the maximum cost is controlled by the coarsest level, and if $\beta < \gamma$, the finest level governs the dominant cost. Finally, when $\beta = \gamma$, the cost at each level is roughly evenly distributed.

Furthermore, it is clear from the first relation in Eq. (37) that as $l \to \infty$,

$$\frac{|\mathbb{E}(u_{h_l}) - \mathbb{E}(u)|}{\sqrt{\lambda_m}} \to 0.$$

However, $\mathbb{E}(u)$ is analytical and not known. Therefore, the deterministic error is defined via the triangle inequality as [9]

$$\frac{|\mathbb{E}(u_{h_l}) - \mathbb{E}(u_{h_{l-1}})|}{\sqrt{\lambda_m}} \leq c_\alpha h_l^\alpha. \tag{39}$$

### 4.2. Multilevel Monte Carlo estimation of variance

To enhance the characterization of the probabilistic system response $u_{h_L}(\boldsymbol{x}, \omega)$, this paper also focuses on defining the MLMC estimator of $\mathbb{V}\mathrm{ar}(u_{h_L})$ using h-statistics, expressed in the form [35]:

$$\widehat{\mathrm{h}}_2^{\mathrm{ML}}(u_{h_L,\{N_l\}}) := \widehat{\mathrm{h}}_2^{\mathrm{MC}}(u_{h_0,N_0}) + \sum_{l=1}^{L}\left(\widehat{\mathrm{h}}_2^{\mathrm{MC}}(u_{h_l,N_l}) - \widehat{\mathrm{h}}_2^{\mathrm{MC}}(u_{h_{l-1},N_l})\right). \tag{40}$$

Here, $\widehat{\mathrm{h}}_2^{\mathrm{MC}}(u_{h_0,N_0})$ is the MC estimator of $\mathbb{V}\mathrm{ar}(u_{h_0})$ with $N_0$ samples; $\widehat{\mathrm{h}}_2^{\mathrm{MC}}(u_{h_l,N_l})$ and $\widehat{\mathrm{h}}_2^{\mathrm{MC}}(u_{h_{l-1},N_l})$ represent the MC estimation of $\mathbb{V}\mathrm{ar}(u_{h_l})$ and $\mathbb{V}\mathrm{ar}(u_{h_{l-1}})$ using $N_l$ samples, respectively. For simplification, we introduce:

$$Z_l = \begin{cases} \widehat{\mathrm{h}}_2^{\mathrm{MC}}(u_{h_0,N_0}), & l = 0,\\ \widehat{\mathrm{h}}_2^{\mathrm{MC}}(u_{h_l,N_l}) - \widehat{\mathrm{h}}_2^{\mathrm{MC}}(u_{h_{l-1},N_l}), & l > 0. \end{cases} \tag{41}$$

Note that, for $l > 0$, $u_{h_l,N_l}$ and $u_{h_{l-1},N_l}$ in $Z_l$ are determined using the same random seed. Thus, the expansion in Eq. (40) is rewritten as

$$\widehat{\mathrm{h}}_2^{\mathrm{ML}} = \sum_{l=0}^{L} Z_l. \tag{42}$$

Similar to MLMC mean estimation, MLMC variance estimator is also obtained as the telescopic sum of the difference of MC variance estimates on the coarser levels. Therefore, in correspondence to Eq. (16), the MSE of the multilevel estimator $\widehat{\mathrm{h}}_2^{\mathrm{ML}}$ takes the form:

$$\mathrm{MSE}(\widehat{\mathrm{h}}_2^{\mathrm{ML}}) = \mathbb{V}\mathrm{ar}\left(\widehat{\mathrm{h}}_2^{\mathrm{ML}}\right) + (\mathbb{V}\mathrm{ar}(u_{h_L}) - \mathbb{V}\mathrm{ar}(u))^2. \tag{43}$$

Under further consideration that the quantity $Z_l$, for $l \geq 0$, is sampled independently, one may express the first term in the above equation as

$$\mathbb{V}\mathrm{ar}\left(\widehat{\mathrm{h}}_2^{\mathrm{ML}}\right) = \sum_{l=0}^{L}\mathbb{V}\mathrm{ar}(Z_l), \tag{44}$$

in which the variance $\mathbb{V}\mathrm{ar}(Z_l)$ is further defined—similar to Eq. (18)—by $\mathbb{V}_{l,2}/N_l$. Therefore, Eq. (43) is reformulated as

$$\mathrm{MSE}(\widehat{\mathrm{h}}_2^{\mathrm{ML}}) = \sum_{l=0}^{L}\frac{\mathbb{V}_{l,2}}{N_l} + (\mathbb{V}\mathrm{ar}(u_{h_L}) - \mathbb{V}\mathrm{ar}(u))^2. \tag{45}$$

Analogous to the MSE of MC variance estimator in Eq. (20), here, the MSE is also split into statistical error—which is of the order

$\mathcal{O}(N_l^{-1})$—and square of deterministic bias. Furthermore, the quantity $\mathbb{V}_{l,2}$ is analytical, and the corresponding unbiased MC estimation, denoted $\widehat{\mathbb{V}}^{\mathrm{MC}}_{l,2}$, is detailed in Appendix D.

**Scale-invariant error estimator for MLMC variance estimation**: The MSE of estimator $\widehat{\mathrm{h}}^{\mathrm{ML}}_2$ in Eq. (45) is scale-dependent. Therefore, the MSE of multilevel variance estimator is transformed to a scale-invariant version $\mathrm{NMSE}(\widehat{\mathrm{h}}^{\mathrm{ML}}_2)$ by considering $\lambda_v = \mathbb{V}_2(u_{h_L})$ as the normalization entity—as described in Section 3.2. Since, $\mathbb{V}_2(u_{h_L})$ is unknown, we make the assumption that the entity $\mathbb{V}_2(u_{h_l})$ remains approximately close for all values of $l$. That is, the value of $\mathbb{V}_2(u_{h_L})$ on finest level $L$ is replaced by $\mathbb{V}_2(u_{h_0})$ of the coarsest mesh level $l = 0$. Finally, with the unbiased MC estimation of the normalizer $\widehat{\lambda}_v := \widehat{\mathbb{V}}^{\mathrm{MC}}_2(u_{h_0,N_0})$, detailed in Section 3.2 and Appendix B, the new normalized MSE is given in the form:

$$\mathrm{NMSE}(\widehat{\mathrm{h}}^{\mathrm{ML}}_2) = \frac{1}{\widehat{\mathbb{V}}^{\mathrm{MC}}_2}\left(\sum_{l=0}^{L}\frac{\widehat{\mathbb{V}}^{\mathrm{MC}}_{l,2}}{N_l}\right) + \frac{(\mathbb{V}\mathrm{ar}(u_{h_L}) - \mathbb{V}\mathrm{ar}(u))^2}{\widehat{\mathbb{V}}^{\mathrm{MC}}_2}. \tag{46}$$

Here, the first term defines the scale-invariant sampling error and the second term is the normalized squared deterministic error. The accomplishment of a total NMSE $\mathring{\epsilon}^2$—as mentioned in Section 4.1—is justified by ensuring that both errors are less than $\mathring{\epsilon}^2/2$.

One may determine the samples $N_l$ analogously to Eqs. (34)-(36), such that Eq. (36) is substituted in Eq. (35):

$$N_l = \frac{2}{\mathring{\epsilon}^2}\sum_{l=0}^{L}\left(\frac{\widehat{\mathbb{V}}^{\mathrm{MC}}_{l,2}C_l}{\widehat{\mathbb{V}}^{\mathrm{MC}}_2}\right)^{\frac{1}{2}}\left(\frac{\widehat{\mathbb{V}}^{\mathrm{MC}}_{l,2}}{\widehat{\mathbb{V}}^{\mathrm{MC}}_2 C_l}\right)^{\frac{1}{2}}. \tag{47}$$

Note that, here $C_l$ is the computational cost of evaluating one MC sample of $u_{h_l,N_l}$ and $u_{h_{l-1},N_l}$ in the difference term $Z_l$, which is equivalent to determining the cost of $Y_l$ in Eq. (34).

Following the computational complexity of the scale-invariant MLMC mean in Proposition. 4.1, the computing cost of the normalized MLMC variance estimate, similar to the conventional MLMC variance (see [4,35]), is detailed below.

**Proposition 4.2.** *Let us introduce the positive constants $c_\alpha, c_\beta, c_\gamma, \alpha, \beta, \gamma$, such that $\alpha \geq \frac{1}{2}\min(\beta,\gamma)$. Then, one may define the following error bounds, such that*

(*i*) *The deterministic error is bounded by* $\dfrac{|\mathbb{V}ar(u_{h_l}) - \mathbb{V}ar(u)|}{\sqrt{\lambda_v}} \leq c_\alpha h_l^\alpha,$

(*ii*) *The variance* $\widehat{\mathbb{V}}^{MC}_{l,2}$ *decays as* $\dfrac{\widehat{\mathbb{V}}^{MC}_{l,2}}{\lambda_v} \leq c_\beta h_l^\beta,$ (48)

(*iii*) *The computational cost to determine a single realization of $Z_l$ is given as*

$C(Z_l) \leq c_\gamma h_l^{-\gamma}.$

*Thus, one may state that, for any $0 < \mathring{\epsilon} < e^{-1}$, the total normalized mean square error of multilevel Monte Carlo (MLMC) variance estimator is bound by $\mathrm{NMSE}(\widehat{\mathrm{h}}^{ML}_2) < \mathring{\epsilon}^2$. Finally, the computational cost of MLMC variance estimation reads*

$$C(\widehat{\mathrm{h}}^{ML}_2) \leq \begin{cases} c\,\mathring{\epsilon}^{-2}, & \beta > \gamma, \\ c\,\mathring{\epsilon}^{-2}(\log\mathring{\epsilon})^2, & \beta = \gamma, \\ c\,\mathring{\epsilon}^{-2-(\gamma-\beta)/\alpha}, & \beta < \gamma. \end{cases} \tag{49}$$

*Here, $c > 0$ is a positive constant.*

Notably, if the deterministic and stochastic convergences in Eqs. (37) and (48) are correspondingly equal, then the computational complexity of the MLMC variance estimator will be asymptotically equal to that of the mean estimate. However, this depends on the regularity of solution $u(x,\omega)$, as proved by authors in [4]. Otherwise, generally, for a given total NMSE $\mathring{\epsilon}^2$, the multilevel mean estimator runs faster as compared to the variance (elaborated with a numerical example in Section 6).

Furthermore, as noted in Section 4.1, mesh level $l = 0$ contributes the most cost in the first scenario in the Eq. (49); the second case imposes an almost equal cost on all mesh levels, and the final relation demonstrates that the finest level $L$ is the most dominant.

Also, corresponding to Eq. (48), when $l \to \infty$, there is a monotonic decay in $|\mathbb{V}\mathrm{ar}(u_{h_l}) - \mathbb{V}\mathrm{ar}(u)|/\sqrt{\lambda_v}$. Thereby, analogous to Eq. (39), the following condition

$$\frac{|\mathbb{V}\mathrm{ar}(u_{h_l}) - \mathbb{V}\mathrm{ar}(u_{h_{l-1}})|}{\sqrt{\lambda_v}} \leq c_\alpha h_l^\alpha, \tag{50}$$

holds.

## 5. Application to linear elasticity: a model problem

To test the scale-invariant MC and MLMC methods, we consider a framework of linear elliptic PDEs with random coefficients. As a representative example, we focus on linear elasticity as a model problem, where the elasticity tensor is modelled as a matrix-valued random field capturing both the heterogeneous and randomly anisotropic nature of the material.

### 5.1. Deterministic setting

Let $\mathcal{G} \subset \mathbb{R}^d$ be a $d$-dimensional geometry with smooth Lipschitz boundary $\Gamma$. The aim is to determine the displacement vector $\boldsymbol{u} \in \mathbb{R}^d$ (which belongs to the Hilbert space $\mathcal{U}$) at a spatial point $\boldsymbol{x} \in \mathcal{G}$ that completely satisfies the equilibrium equations [42]

$$\begin{aligned} -\mathrm{div}\,\boldsymbol{\sigma}(\boldsymbol{x}) &= \boldsymbol{f}(\boldsymbol{x}), \quad \forall \boldsymbol{x} \in \mathcal{G}, \\ \boldsymbol{u}(\boldsymbol{x}) &= \boldsymbol{u}_0 = \boldsymbol{0}, \quad \forall \boldsymbol{x} \in \Gamma_D, \\ \boldsymbol{\sigma}(\boldsymbol{x})\cdot\boldsymbol{n}(\boldsymbol{x}) &= \boldsymbol{t}(\boldsymbol{x}), \quad \forall \boldsymbol{x} \in \Gamma_N, \end{aligned} \tag{51}$$

describing the linear-elastic behaviour. Here, $\boldsymbol{\sigma}(\boldsymbol{x})$ is the Cauchy stress tensor, which belongs to the space of second-order symmetric tensors $\mathrm{Sym}(d) := \{\boldsymbol{\sigma} \in (\mathbb{R}^d \otimes \mathbb{R}^d) \mid \boldsymbol{\sigma} = \boldsymbol{\sigma}^T\}$; $\boldsymbol{f}(\boldsymbol{x}) \in \mathbb{R}^d$ is the body force; $\boldsymbol{t}(\boldsymbol{x}) \in \mathbb{R}^d$ represents the surface tension on the Neumann boundary $\Gamma_N \subset \Gamma$, and $\boldsymbol{n}(\boldsymbol{x}) \in \mathbb{R}^d$ is the outward unit normal to $\Gamma_N$. For simplicity, a homogeneous boundary condition $\boldsymbol{u}_0 = \boldsymbol{0} \in \mathbb{R}^d$ is considered on the Dirichlet boundary $\Gamma_D \subset \Gamma$. It is also possible to assume that $\Gamma_D \cap \Gamma_N = 0$.

One may further describe the strain–displacement relation as

$$\boldsymbol{\varepsilon}(\boldsymbol{x}) = \frac{1}{2}\left(\nabla\boldsymbol{u} + \nabla\boldsymbol{u}^T\right), \quad \forall \boldsymbol{x} \in \mathcal{G}, \tag{52}$$

where $\boldsymbol{\varepsilon}(\boldsymbol{x}) \in \mathrm{Sym}(d)$ denotes an infinitesimal second-order symmetric strain tensor. Finally, the material constitutive equation is of the linear form

$$\boldsymbol{\sigma}(\boldsymbol{x}) = \boldsymbol{C}(\boldsymbol{x}) : \boldsymbol{\varepsilon}(\boldsymbol{x}), \quad \forall \boldsymbol{x} \in \mathcal{G}, \tag{53}$$

in which $\boldsymbol{C}(\boldsymbol{x})$ represents a spatially varying fourth-order positive-definite symmetric elasticity tensor. Here, the notion of symmetry signifies the major ($C_{ijkl} = C_{klij}$) and minor symmetries ($C_{ijkl} = C_{jikl} = C_{ijlk}$). As a result, one can map a $\boldsymbol{C}(\boldsymbol{x})$ tensor to a second-order tensor $C(\boldsymbol{x})$, more generally known as the elasticity matrix. The reduced $C(\boldsymbol{x})$ matrix belongs to a family of $(n \times n)$, where $n = d(d+1)/2$, real-valued positive-definite symmetric matrices:

$$\mathrm{Sym}^+(n) = \{C \in (\mathbb{R}^n \otimes \mathbb{R}^n) \mid C = C^T, \boldsymbol{z}^T C \boldsymbol{z} > \boldsymbol{0}, \forall \boldsymbol{z} \in \mathbb{R}^n \setminus \boldsymbol{0}\}. \tag{54}$$

Accordingly, conforming to Voigt notation, Eq. (53) transforms to $\sigma(\boldsymbol{x}) = C(\boldsymbol{x}) \cdot \varepsilon(\boldsymbol{x})$, with $\sigma(\boldsymbol{x}) \in \mathbb{R}^d$ and $\varepsilon(\boldsymbol{x}) \in \mathbb{R}^d$ denoting the stress and strain vectors, respectively.

### 5.2. Stochastic modelling of material uncertainty: a reduced parametric approach

Material law as described previously is complicated when highly heterogeneous and anisotropic materials—such as bone tissue, see

Section 6—are to be modelled. To include aleatoric uncertainty with heterogeneity in Eq. (53), the material properties have to be modelled as random and spatially dependent [49,60]. In this paper, a probabilistic point of view is studied, in which the $C(\boldsymbol{x})$ matrix is modelled as a matrix-valued second-order random field $C(\boldsymbol{x},\omega)$ on a probability space $(\Omega,\mathcal{F},\mathbb{P})$. In other words, the random elasticity matrix can be modelled as a mapping:

$$C(\boldsymbol{x},\omega) : \mathcal{G}\times\Omega \to \mathrm{Sym}^{+}(n). \tag{55}$$

Furthermore, in practise, the usual assumption is that the elasticity matrix $C(\boldsymbol{x})$ globally follows a certain type of symmetry (e.g., isotropic, orthotropic, etc.; see [6,11] for invariance classification of elasticity matrix $C$), even though the experiments today cannot provide this information with full certainty. Therefore, to include uncertainty in the type of material symmetry, in this paper, we model the elasticity matrix $C(\boldsymbol{x})$ as per the reduced parametric approach (also commonly known as the non-parametric method) [23,61,62]. That is, the random matrix $C(\boldsymbol{x},\omega)$ follows a specific type of symmetry only in the mean, whereas each of the realisations belongs to the triclinic system, which is the lowest order of symmetry for elasticity-type tensors. This gives our model a full degree of freedom in a case where the predefined mean symmetry turns out to be incorrect; one would still be able to identify other types of invariances given experimental data. This is, however, not the case if we assume that each of the realisations is constrained. If the specific symmetry class of the material—beyond the triclinic case—is known for both the ensemble (population) and its mean, one may employ the stochastic modelling frameworks presented in [39], and [58,59], which also has the ability to separate the modelling of strength, eigen-strain distribution, and spatial orientation, allowing for independent control of each component; however, this remains beyond the scope of the present study.

Following this, we model the homogeneous mean matrix as $\mathbb{E}(C(\boldsymbol{x},\omega)) = U^{T}U$, in which the term $U \in \mathbb{R}^{n\times n}$ represents an upper triangular matrix—this square-type (Cholesky) factorization ensures the positive-definiteness of the mean matrix. Then, to allow uncertainty into the model, the mean formulation is extended to

$$C(\boldsymbol{x},\omega) = U^{T}T(\boldsymbol{x},\omega)U, \tag{56}$$

such that

$$\mathbb{E}(T(\boldsymbol{x},\omega)) = I \tag{57}$$

holds. Here, $T(\boldsymbol{x},\omega)$ is a matrix-valued random variable that also resides in $\mathrm{Sym}^{+}(n)$, the mean value of which is an identity matrix $I \in \mathrm{Sym}^{+}(n)$. In this manner, the mean behaviour is controlled by $\mathbb{E}(C(\boldsymbol{x},\omega))$, and the random fluctuations are governed by $T(\boldsymbol{x},\omega)$. To construct a random ensemble $T(\boldsymbol{x},\omega)$, one may use the maximum entropy optimisation principle under the constraints mentioned in the previous list (but with a mean defined in Eq. (57)). Henceforth, $T(\boldsymbol{x},\omega)$ can further be factorised as

$$T(\boldsymbol{x},\omega) = V(\boldsymbol{x},\omega)^{T}V(\boldsymbol{x},\omega), \tag{58}$$

and by substitution in Eq. (56), one obtains

$$C(\boldsymbol{x},\omega) = U^{T}V(\boldsymbol{x},\omega)^{T}V(\boldsymbol{x},\omega)U. \tag{59}$$

Here, $V(\boldsymbol{x},\omega) \in \mathbb{R}^{n\times n}$ denotes the upper triangular random matrix with entries

$$V_{ij} = \begin{cases} \frac{\delta_T}{\sqrt{d+1}}\theta_{ij}(\boldsymbol{x},\omega), \text{if } i<j \\ \frac{\delta_T}{\sqrt{d+1}}\sqrt{N_c\Gamma_{\alpha_j}(\boldsymbol{x},\omega)}, \text{if } i=j, \end{cases} \tag{60}$$

given that the elements $V_{ij}, i \le j$ are independent. Observe that the non-diagonal upper triangular elements are modelled as independent Gaussian random fields $\theta_{ij}(\boldsymbol{x},\omega), 1 \le i \le j \le n$ with zero mean and unit variance. On the other hand, the diagonal entries are gamma-distributed positive-definite random fields:

$$\Gamma_{\alpha_j}(\boldsymbol{x},\omega) = F^{-1}_{\Gamma_\alpha} \circ \mathrm{erf}(\theta_{ij}(\boldsymbol{x},\omega)). \tag{61}$$

Here,

$$\alpha_j = (d+1)/(2\delta_T^2) + (1-j)/2 \tag{62}$$

is a positive real number; erf is the standard Gaussian distribution function; and $F^{-1}_{\Gamma_\alpha}$ is the inverse gamma cumulative distribution function. This assures that the diagonal elements are strictly positive, and therefore, the random matrix $T(\boldsymbol{x},\omega)$ also remains positive-definite. Reverting to the above description in Eq. (60), where $N_c$ is a normalization constant of the resultant function (given in Eq. (61)), the value of which equals 2. Furthermore, $\delta_T := [0,1] \in \mathbb{R}$ defined as Eq. (63), is a scalar value that controls the dispersion of $T(\boldsymbol{x},\omega)$:

$$\delta_T = \left\{ \frac{1}{d}\mathbb{E}\left[ \| T(\boldsymbol{x},\omega) - I \|^2 \right] \right\}^{1/2}. \tag{63}$$

The coefficient of dispersion parameter $\delta_T$ is chosen such that

$$\delta_C = \left[ \frac{\mathbb{E}\left\{ \|C(\boldsymbol{x},\omega) - \mathbb{E}(C(\boldsymbol{x},\omega))\|^2 \right\}}{\|\mathbb{E}(C(\boldsymbol{x},\omega))\|^2} \right]^{1/2} = \frac{\delta_T}{\sqrt{d+1}} \left[ 1 + \frac{(\mathrm{tr}(\mathbb{E}(C(\boldsymbol{x},\omega))))^2}{\mathrm{tr}(\mathbb{E}(C(\boldsymbol{x},\omega)))^2} \right]^{1/2} \tag{64}$$

holds. Here, $\delta_C := [0,1] \in \mathbb{R}$ is the coefficient of dispersion of the random matrix $C(\boldsymbol{x},\omega)$.

### 5.3. Stochastic setting

The description of the random elasticity matrix field $C(\boldsymbol{x},\omega)$ leads to the apparent transformation of a linear-elastic material model to a stochastic model. The aim is to determine the random displacement vector field $\boldsymbol{u}(\boldsymbol{x},\omega) : \mathcal{G}\times\Omega \to \mathbb{R}^d$. Therefore, the equilibrium equations are rewritten in the form:

$$\begin{aligned} -\mathrm{div}\,\boldsymbol{\sigma}(\boldsymbol{x},\omega) &= \boldsymbol{f}(\boldsymbol{x}), \quad \forall \boldsymbol{x}\in\mathcal{G}, \omega\in\Omega, \\ \boldsymbol{u}(\boldsymbol{x},\omega) &= \boldsymbol{u}_0 = \boldsymbol{0}, \quad \forall \boldsymbol{x}\in\Gamma_D, \omega\in\Omega, \\ \boldsymbol{\sigma}(\boldsymbol{x},\omega)\cdot\boldsymbol{n}(\boldsymbol{x}) &= \boldsymbol{t}(\boldsymbol{x}), \quad \forall \boldsymbol{x}\in\Gamma_N, \omega\in\Omega, \end{aligned} \tag{65}$$

in which the boundary conditions and body forces $\boldsymbol{f}(\boldsymbol{x})$ remain deterministic. Further, the linearized kinematics relationship is transformed to

$$\boldsymbol{\varepsilon}(\boldsymbol{x},\omega) = \frac{1}{2}\left( \nabla\boldsymbol{u}(\boldsymbol{x},\omega) + \nabla\boldsymbol{u}(\boldsymbol{x},\omega)^{T} \right), \forall \boldsymbol{x}\in\mathcal{G}, \omega\in\Omega, \tag{66}$$

where the $\nabla(\cdot)$ operator is taken in a weak sense. Finally, the constitutive law is represented by

$$\boldsymbol{\sigma}(\boldsymbol{x},\omega) = \boldsymbol{C}(\boldsymbol{x},\omega) : \boldsymbol{\varepsilon}(\boldsymbol{x},\omega), \forall \boldsymbol{x}\in\mathcal{G}, \omega\in\Omega. \tag{67}$$

By carrying out the variational formulation of the above stochastic partial differential equations on $\mathcal{G}$ and further discretizing in a finite element setting [70], one searches for the solution $\boldsymbol{u}_h(\boldsymbol{x},\omega) : \mathcal{U}_h\times\Omega \to \mathbb{R}^d$ in a finite subspace, $\mathcal{U}_h$, as described in [30,38,52].

In computational stochastic mechanics, a vast body of literature exists on numerical methods for obtaining the stochastic solution $\boldsymbol{u}_h(\boldsymbol{x},\omega)$ after semi-discretization. Some well-known approaches that fall into the category of series–expansion methods include the spectral stochastic finite element method [19,43,44], the perturbation method [28,33,36], and the Neumann expansion method [55,68]. The present study, however, focuses on another class of techniques that directly integrate the statistics of the response, of which, Monte Carlo simulation [50,56]

has traditionally been used for assessing the validity of other methods. Further reviews and discussions of the procedures can be found in [19,43,54,64].

The goal of this study is to determine the total displacement scalar field in Euclidean norm, i.e., $u_h^{(t)}(\boldsymbol{x},\omega) = \|\boldsymbol{u}_h(\boldsymbol{x},\omega)\|$, and estimate the second-order statistics such as the mean and variance of sampled response $u_{h,N}^{(t)} := [u_h^{(t)}(\boldsymbol{x},\omega_i)]_{i=1}^N$ using the scale-invariant MC and MLMC as detailed in Sections (3) and (4). Note that a comparison of the proposed MC and MLMC techniques with the previously listed series–expansion approaches lies outside the scope of this work.

## 6. Numerical results: 2D human femur

In this section, we examine the performance of the scale-invariant MC and MLMC algorithms on a linear elastic material model of a 2D human femoral bone, which is a highly heterogeneous and anisotropic material.

### 6.1. Specifications

A two-dimensional proximal femur bone geometry with a body width of approximately 7 cm and 21.7 cm in total height is considered. Fig. 1 shows the boundary conditions where an in-plane uniform pressure load with a resultant load of 1500 N is applied on top of the bone and zero displacements are considered at the bottom [69]. The finite element method (FEM) based spatial discretization is performed using four-noded plane stress elements. By sampling the probabilistic space, each deterministic simulation is thus executed by the finite element MATLAB-based software Plaston [52], where preconditioned conjugate gradient (PCG) method is used as an iterative solver.

To implement the scale-invariant MLMC method, a sequence of four nested meshes with element size $h_{l-1} = 2h_l$ is considered, as shown in Fig. 2. One may notice the identical implementation of boundary conditions across all the mesh levels. The corresponding mesh specifications are tabulated in Table 1.

Femoral bone tissue is not only a highly heterogeneous material but also anisotropic, with its precise type of elastic symmetry remaining uncertain [31,53]. To account for this, we model the material stiffness using a matrix-valued random field $C(\boldsymbol{x},\omega)$, as defined in Section 5.2. We consider the average orthotropic elastic properties of human cortical femoral bone, as reported in the experimental study conducted on 60 specimens by [1]. The corresponding elastic coefficients, restricted to 2D, are listed in Table 2, and are used to define the homogeneous mean matrix $\mathbb{E}(C(\boldsymbol{x},\omega))$ of the random elasticity tensor field $C(\boldsymbol{x},\omega)$. By setting the coefficient of dispersion of the random elasticity matrix field to $\delta_C = 0.1$, the coefficient of dispersion of the matrix-valued random field $T(\boldsymbol{x},\omega)$ is determined using Eq. (64).

Accordingly, the fluctuation matrix $T(\boldsymbol{x},\omega)$ is modelled as a nonlinear transformation of 6 (as $n = 3,\ d = 2$) independent scalar Gaussian random fields. Each of these is approximated via a truncated (up to $M$ terms) Kosambi-Karhunen-Loève expansion [29,34,37], expressed as:

$$\theta_{ij}(\boldsymbol{x},\omega) = \overline{\theta}_{ij}(\boldsymbol{x}) + \sum_{k=1}^{M} \sqrt{\eta_k}\psi_k(\boldsymbol{x})\xi_k(\omega). \tag{68}$$

Here, $\overline{\theta}_{ij}(\boldsymbol{x})$ denotes the spatially varying mean field, $\xi_k(\omega)$ are the mutually uncorrelated and independent standard Gaussian random variables, and $(\eta_k,\psi_k)$ are the eigenpairs of the covariance operator associated with the autocorrelation kernel $\{R_{ij}(r) : R_{ij}(0) = 1, 1 \le i \le j \le n\}$. These eigenpairs are obtained via a finite element discretization of the Fredholm integral equation of the second kind [3,30,45,63]. In this work, we assume a Gaussian-type covariance structure:

$$R_{ij}(r) = \varrho^2 \exp(-l_c^{-2} r^2), \tag{69}$$

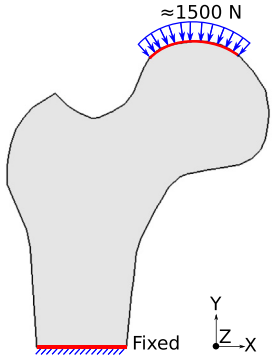


**Fig. 1.** Geometry and boundary conditions.

**Table 1**
Mesh specifications of 2D meshes.

| $l$ | Elements | Nodes | DOF |
|---|---|---|---|
| 0 | 171 | 206 | 396 |
| 1 | 684 | 753 | 1476 |
| 2 | 2736 | 2873 | 5688 |
| 3 | 10,944 | 11,217 | 22,320 |

**Table 2**
Orthotropic material parameters.

| Young's modulus (GPa) | Poisson's ratio | Shear modulus (GPa) |
|---|---|---|
| $E_1 = 12$<br>$E_2 = 20$ | $\nu_{21} = 0.371$ | $G_{12} = 5.61$ |

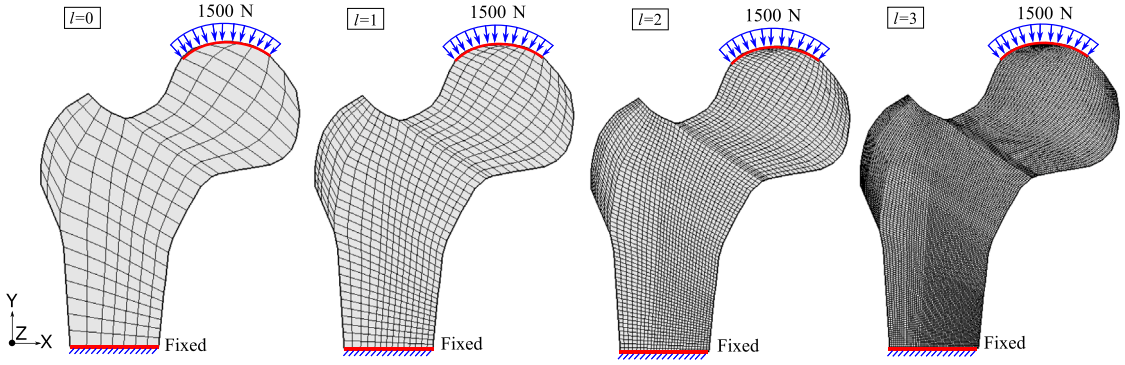


**Fig. 2.** Nested mesh levels of 2D femur bone.

in which, $r = \|\boldsymbol{x}_1 - \boldsymbol{x}_2\|$ is the Euclidean distance between spatial points $(\boldsymbol{x}_1, \boldsymbol{x}_2) \in \mathcal{G}$ with $r \geq 0$, and $\varrho^2$ is the marginal variance. Moreover, each of $R_{ij}$ are individually parametrized by a vector of correlation lengths $l_c \in \mathbb{R}^d_+$. For simplicity, we adopt identical correlation lengths across all components of autocorrelation $R_{ij}$. Note that the standard Gaussian random variables $\xi_k(\omega)$ in Eq. (68) are generated using MATLAB's built-in *randn* function, which by default employs the ziggurat algorithm [40,41].

In this study, the mean $\overline{\theta}_{ij}(\boldsymbol{x})$ is set to be spatially constant with a value of zero and the variance $\varrho^2 = 1$. Under the assumptions considered for constructing the MLMC estimator in Section 4.1, each matrix-valued random field $C(\boldsymbol{x}, \omega)$ and the corresponding Gaussian fields $\theta_{ij}(\boldsymbol{x}, \omega)$ are modelled independently on each mesh level $l$. Whereas, for the definition of the difference terms $Y_l$ and $Z_l$ in Eqs. (30) and (41), the random field on the coarse mesh level $l-1$ is obtained by directly mapping it from the fine level $l$ at the intersecting common spatial nodes. Furthermore, the Gaussian autocorrelation function in Eq. (69) is defined with a correlation length of 3.5 cm in both $x$- and $y$-directions, across all four levels of meshes. The expansion, given in Eq. (68), is truncated to $M = 100$ terms on all mesh levels, chosen based on the decay of eigenvalues $\eta_k$ on fine mesh $L$, as illustrated in Fig. 3. For further optimality, one may also use level-dependent truncations, as discussed in [65].

The random anisotropy of the material at a fixed spatial location $\boldsymbol{x}$ on the coarse mesh level $l = 0$, modelled by the matrix-valued random field $C(\cdot, \omega)$, is illustrated in Fig. 4. Here we demonstrate the characteristic

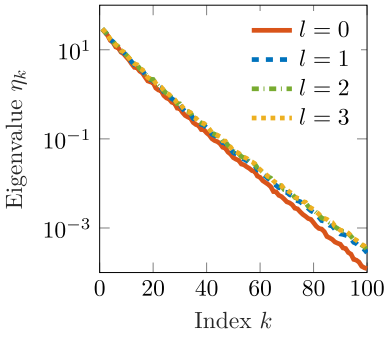


**Fig. 3.** Decay of eigenvalues across all mesh levels.

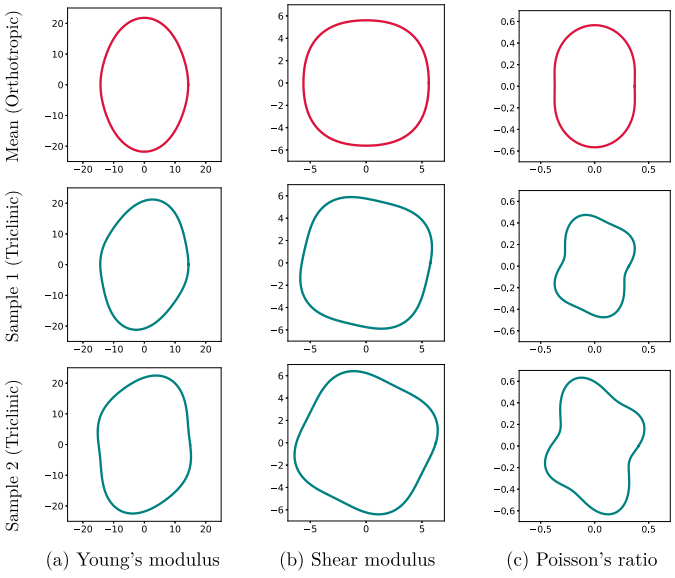


**Fig. 4.** Visualization of random anisotropy $C(\cdot, \omega)$ on mesh $l = 0$.

directional elastic parameters in the columns—namely, Young's modulus, shear modulus, and Poisson's ratio—using the open-source software ELATE [17]. The first row of the figure presents the orthotropic elastic properties corresponding to the mean matrix $\mathbb{E}(C(\boldsymbol{x}, \omega))$, while the subsequent rows display triclinic characteristics of two individual realizations, $C(\cdot, \omega_1)$ and $C(\cdot, \omega_2)$, where one may notice the variation in shape and size of all three parameters as compared to the mean. Additionally, we visualize the spatial variation of a single realization of the random field component $C_{1,1}(\boldsymbol{x}, \cdot)$ across all mesh levels in Fig. 5.

The objective of the study is to compute the MLMC mean $\widehat{\mu}^{\mathrm{ML}}$ and variance $\widehat{\mathrm{h}}_2^{\mathrm{ML}}$ estimates of the total displacement random field $u^{(t)}_{h_L}(\boldsymbol{x}, \omega)$. The procedure for implementation of the scale-invariant MLMC is similar to that of the conventional MLMC method; the primary difference is in the usage of normalized error instead of absolute error; see [21] for the algorithm. In this study, we assume that the optimal finest level (here, $L = 3$) is known i.e., the normalized squared discretization errors, from Eqs. (33) and (46), are less than $\mathring{\epsilon}^2/2$. Therefore, the focus is only on satisfying the condition of normalized sampling errors; in other words, the objective is to ensure that the normalized sampling errors in Eqs. (33) and (46) are less than $\mathring{\epsilon}^2/2$. Furthermore, to avoid the interpolation error, the terms $Y_l$ and $Z_l$ in Eqs. (30) and (41) are calculated only at finite element nodes that have the same common spatial coordinates between all four levels of meshes. This further means that the MLMC mean and variance estimates of the system response on the finest level $L$ are evaluated only at these common nodes.

### 6.2. Screening test

An a priori performance analysis of MLMC mean and variance estimators is conducted using the so-called screening test, in which a fixed number of 50 samples is considered over four levels of meshes. For the MLMC mean estimate, we assume—for simplicity—that the estimated normalizing function $\widehat{\mathrm{h}}_2^{\mathrm{MC}}(u^{(t)}_{h_0,N_0})$ (from Eq. (33)) is spatially constant, i.e., $\widehat{\lambda}_m = \max(\widehat{\mathrm{h}}_2^{\mathrm{MC}}(u^{(t)}_{h_0,N_0}))$. With this, by considering a $L_\infty$ norm over the error bounds in Eq. (39) and the second relation in Eq. (37), one obtains:

$$\begin{aligned} \frac{\max|\widehat{\mu}^{\mathrm{MC}}(Y_l)|}{\sqrt{\widehat{\lambda}_m}} &\leq c_\alpha h_l^\alpha, \\ \frac{\max\left(\widehat{\mathrm{h}}_2^{\mathrm{MC}}(Y_l)\right)}{\widehat{\lambda}_m} &\leq c_\beta h_l^\beta. \end{aligned} \tag{70}$$

As to the MLMC variance estimate, similar to the assumptions made in the previous equations, one may consider $\widehat{\lambda}_v = \max(\widehat{\mathbb{V}}_2^{\mathrm{MC}}(u^{(t)}_{h_0,N_0}))$ as the normalization constant. Thereby, Eq. (50) and the second case in Eq. (48) are rewritten as

$$\begin{aligned} \frac{\max|Z_l|}{\sqrt{\widehat{\lambda}_v}} &\leq c_\alpha h_l^\alpha, \\ \frac{\max\left(\widehat{\mathbb{V}}^{\mathrm{MC}}_{l,2}\right)}{\widehat{\lambda}_v} &\leq c_\beta h_l^\beta. \end{aligned} \tag{71}$$

Following this, Fig. 6 shows an overview of the corresponding results. Note that the normalization constants $\widehat{\lambda}_m$ and $\widehat{\lambda}_v$ for the screening test are also set by 50 initial samples; however, during the implementation of the MLMC method, the estimates of both constants are progressively updated as additional samples become available. The top left and right plots show the behaviour of the logarithm of ratios defined in Eqs. (70) and (71). The deterministic decay of difference terms $\max|\widehat{\mu}^{\mathrm{MC}}(Y_l)|$ and $\max(Z_l)$ can be seen. A stochastic convergence of the quantities $\max(\widehat{\mathrm{h}}_2^{\mathrm{MC}}(Y_l))$ and $\max(\widehat{\mathbb{V}}^{\mathrm{MC}}_{l,2})$ is also shown in the second plot. Interestingly, one may notice that both the convergences (deterministic and stochastic) pertaining to mean and variance estimates decay in a

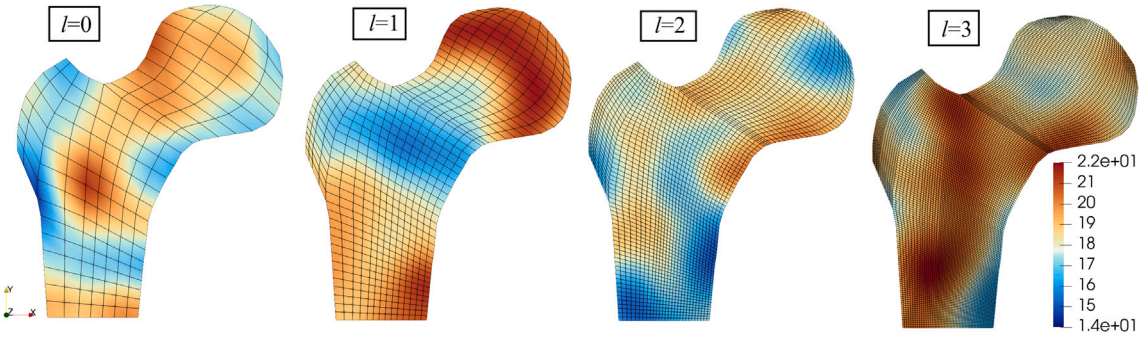


**Fig. 5.** Visualization of spatial variation of a realization $C_{1,1}(\boldsymbol{x},\cdot)$.

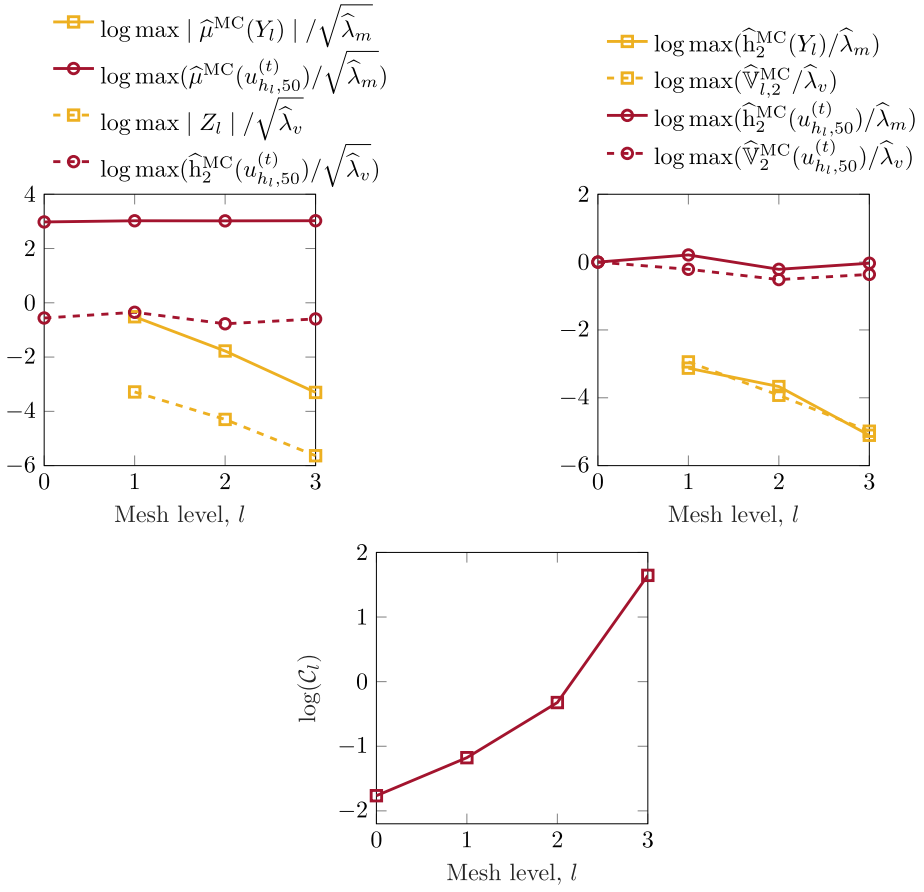


**Fig. 6.** Screening results of scale-invariant MLMC mean and variance estimators.

similar manner. On the other hand, in the top left plot, the quantities $\max(\widehat{\mu}^{\mathrm{MC}}(u^{(t)}_{h_l,50}))$ and $\max(\widehat{\mathrm{h}}_2^{\mathrm{MC}}(u^{(t)}_{h_l,50}))$ stay approximately constant at all values of $l$. Similarly, on the right hand side, the entities $\max(\widehat{\mathrm{h}}_2^{\mathrm{MC}}(u^{(t)}_{h_l,50}))$ and $\max(\widehat{\mathbb{V}}_2^{\mathrm{MC}}(u^{(t)}_{h_l,50}))$ are also approximately constant for varying values of $l$—which meets the assumption for consideration of standardizing factors $\widehat{\lambda}_m$ and $\widehat{\lambda}_v$ on coarse mesh $l = 0$ made in Sections (4.1) and (4.2). Furthermore, the logarithmic computational time of running one sample $C_l$ on each mesh level $l$ is shown at the bottom. The corresponding values are obtained by recording the timings for the 50 considered screening samples on a 2.3 GHz Intel Core i5 processor with 8GB of RAM and taking the average. As the results show, computing becomes more expensive as the mesh refinement increases.

Finally, the decay rates and constants corresponding to Eqs. (70) and (71), as well as the third equation in Eqs. (37) and (48) are evaluated by determining the slopes and y-intercepts of respective logarithmic quantities in Fig. 6, which are further summarised in Table 3. Clearly, all the constants are positive, and the condition $\alpha \geq \frac{1}{2}\min(\beta,\gamma)$ is satisfied, verifying the assumption made in Propositions. 4.1 and 4.2. One may notice the decay rates $\alpha$ and $\beta$ are closer, and equal values of order $\gamma$, for both mean and variance estimators. The constant $c_\beta$ remains close, with equal $c_\gamma$ and differing $c_\alpha$, for both estimates. As $\beta < \gamma$ for both mean and variance estimators, the computational complexity of MLMC estimates follows the third scenario in Eqs. (38) and (49), respectively. Moreover, in the considered example with a fixed number of mesh levels, the total computing cost of both MLMC estimators is dependent on

**Table 3**
Convergence results of MLMC mean and variance estimators.

| Statistic | $\alpha$ | $\beta$ | $\gamma$ | $c_\alpha$ | $c_\beta$ | $c_\gamma$ |
|---|---|---|---|---|---|---|
| Mean | 2.01 | 1.43 | 1.6 | 2.52 | 0.14 | 0.13 |
| Variance | 1.70 | 1.47 | | 0.13 | 0.15 | |

**Table 4**
Number of MC samples on level $l = 3$ with varying stochastic accuracies.

| NMSE, $\mathring{\epsilon}^2/2$ | Number of samples, $N$ |
|---|---|
| $6\times10^{-4}$ | 1667 |
| $4\times10^{-4}$ | 2500 |
| $2\times10^{-4}$ | 5000 |

the stochastic convergence order $\beta$. As the order $\beta$ for variance estimate is slightly higher than that of the mean, for a given sampling accuracy $\mathring{\epsilon}^2/2$, the estimation of MLMC variance is projected to be more expensive than the mean estimation.

### 6.3. Performance analysis

Fig. 7 shows an overview of the performance of scale-invariant MLMC. The top plot shows the propagation of a maximum number of samples $\max(N_l)$ on each level $l$ for both MLMC mean and variance estimators, corresponding to varying normalized sampling accuracies $\mathring{\epsilon}^2/2$. One may observe that $N_l$ decreases monotonically with increasing $l$ for both estimates. For all of the investigated accuracies, it is obvious that the variance estimate requires more samples than the mean estimate. On the other hand, the maximum number of MC samples run on mesh level $l = 3$, satisfying the given stochastic NMSEs, is tabulated in Table 4.

The bottom left plot compares the overall cost of MLMC and MC mean and variance estimation to given normalized sampling errors. Certainly, MLMC estimates have a faster convergence rate than the MC approach. Furthermore, the MC cost of mean and variance exhibits non-asymptotic behaviour. However, the cost of MLMC estimates differs, with the variance estimate being relatively more expensive than the mean—as stated in Section 4.2. As scale-invariant error estimates are used in this investigation, it is possible to conclude that the cost of the MLMC mean and variance estimator is asymptotic.

The MLMC estimators for the mean and variance are observed to be approximately 6 and 4 times more efficient, respectively, than their standard MC counterparts across all levels of normalized sampling accuracy. In other words, approximate cost savings of 83 % and 75 % are reported by the mean and variance of MLMC estimates, respectively, as compared to the MC costs. The bottom-right graphic also shows the scale invariance element of the MLMC convergence for an individual statistic in two distinct units (cm and m). Both the MLMC mean (in cm and m) and variance (in $cm^2$ and $m^2$) convergence remain unchanged.

Finally, the summary of the given and estimated maximum value of stochastic NMSEs of MLMC mean and variance estimators is listed in Table 5. It can be seen that the achieved maximum NMSEs of mean and variance are within the given limits of $\mathring{\epsilon}^2/2$. Furthermore, the maximum absolute sampling accuracies, a product of normalizing constant and maximum sampling NMSE, are also presented in the table. For a given NMSE value, the absolute MSE of variance is much smaller in magnitude as compared to the mean; however, both belong to different scales. This emphasizes the need for normalized error estimates to ensure easier interpretation of performance between the MLMC estimators.

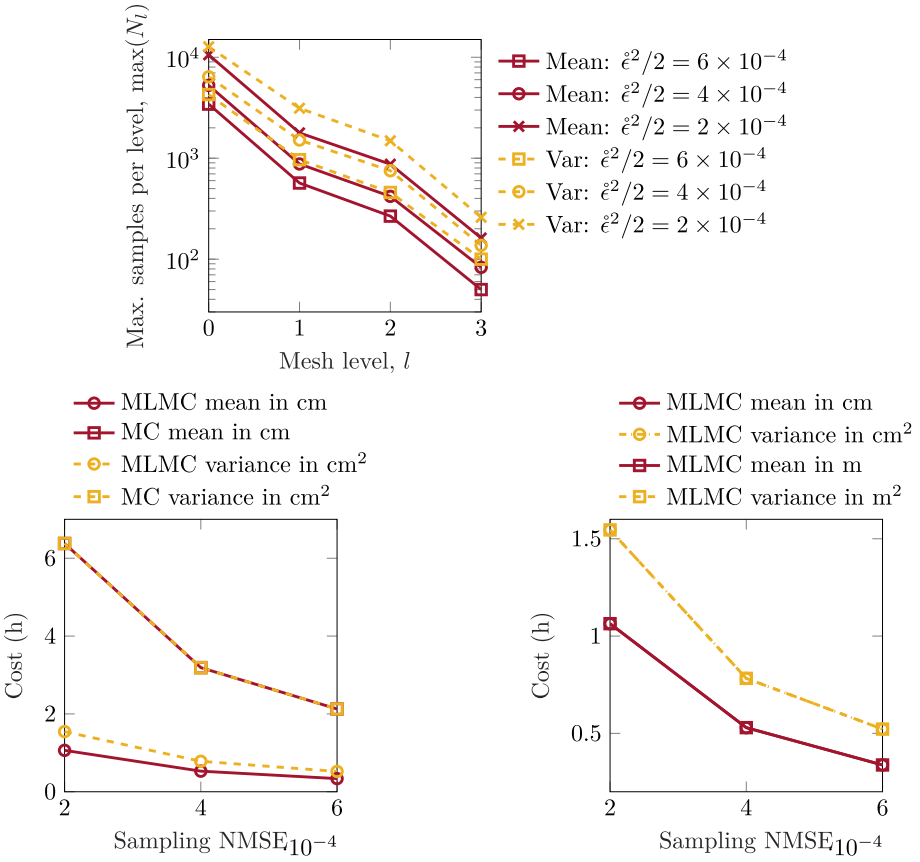


**Fig. 7.** Performance of scale-invariant MLMC mean and variance estimators.

**Table 5**
Given and estimated sampling NMSEs and their corresponding absolute MSEs of MLMC mean and variance estimates.

| Given NMSE, $\hat{\epsilon}^2/2$ | Estimated NMSE | Estimated MSE (cm$^2$) |
|---|---|---|
| (a) Mean | | |
| $2\times10^{-4}$ | $1.97\times10^{-4}$ | $9.459\times10^{-10}$ |
| $4\times10^{-4}$ | $3.99\times10^{-4}$ | $1.889\times10^{-9}$ |
| $6\times10^{-4}$ | $5.98\times10^{-4}$ | $2.859\times10^{-9}$ |
| (b) Variance | | |
| Given NMSE, $\hat{\epsilon}^2/2$ | Estimated NMSE | Estimated MSE (cm$^4$) |
| $2\times10^{-4}$ | $1.98\times10^{-4}$ | $9.387\times10^{-15}$ |
| $4\times10^{-4}$ | $3.89\times10^{-4}$ | $1.856\times10^{-14}$ |
| $6\times10^{-4}$ | $5.83\times10^{-4}$ | $2.789\times10^{-14}$ |

Figs. 8 and 9 compare the mean and variance estimates of total displacement (TD) of the femoral bone between MC and MLMC. For easier interpretation, MC and MLMC results are displayed on a single scale for both mean and variance estimates.

In general, the results are displayed for the normalized mean-squared accuracy of $2 \times 10^{-4}$. Note that, as the displacement values are determined on the finest mesh $L = 3$ at the common nodes corresponding to the coarse mesh $l = 0$, the contour plots are mapped directly on this coarse mesh. In Fig. 8, a maximum mean value of approximately 0.0462 cm can be seen in the region of the pressure load applied. Further, Fig. 9 shows the influence of material uncertainty on the displacement $u^{(t)}_{h_L,\{N_l\}}$ due to the random material model $C(\boldsymbol{x},\omega)$. A maximum variance is also witnessed at the top region of the bone. Furthermore, the right-most plots in both figures clearly demonstrate that the absolute error between the MLMC and standard MC methods is minimal, thereby highlighting the high level of accuracy attained by the MLMC estimators.

## 7. Conclusion

We present novel scale-invariant approaches for estimating the mean and variance of a quantity of interest (QoI) using the multilevel Monte Carlo (MLMC) method. These methods are based on the derivation of normalized mean square error (NMSE) estimates for the classical Monte Carlo (MC) mean and variance estimation. The proposed relative errors are statistically defined using h-statistics, with chosen normalizing factors that are finite and unbiased. Unlike traditional normalization approaches that rely on the squared value of the estimator, which fail to achieve full scale-invariance, the proposed NMSEs are invariant under any linear transformation (scaling and translation) of QoI, and remain robust to changes in its distributional characteristics. Such a standardized formulation reduces interpretational ambiguity and enables a dimensionless assessment of statistical accuracy and computational efficiency across different estimators and scales.

The proposed scale-invariant MC and MLMC methods are tested on a two-dimensional simulation of a human femur modelled as an uncertain linear elastic constitutive law. As bone tissue is a highly heterogeneous and anisotropic material, and that its precise elastic symmetry class is typically unknown or uncertain, the material's elasticity tensor is modelled as a matrix-valued random field. This modelling framework captures both the spatial variability of material properties and random anisotropy by prescribing an elastic symmetry in the mean (e.g., orthotropic) and allowing for triclinic symmetry in individual realizations. The proposed methods then propagate these uncertainties to estimate statistics like the mean and variance of the stochastic total displacement field.

Through normalized error estimates, we compare the computational efficiencies of MLMC and MC estimators for both mean and variance. MLMC significantly outperforms MC in terms of computational cost for both estimates. However, the variance estimate in MLMC requires a

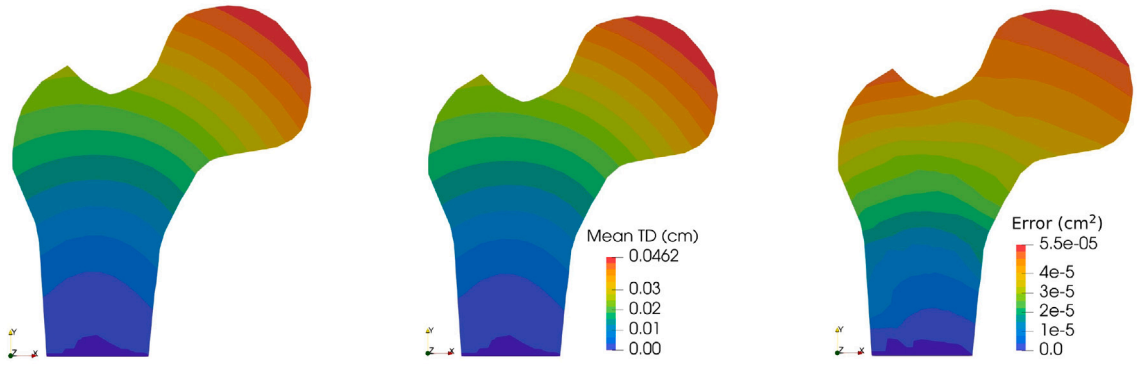


**Fig. 8.** MLMC (left) and MC (middle) mean estimate of total displacement (with identical scale), along with absolute error on the right.

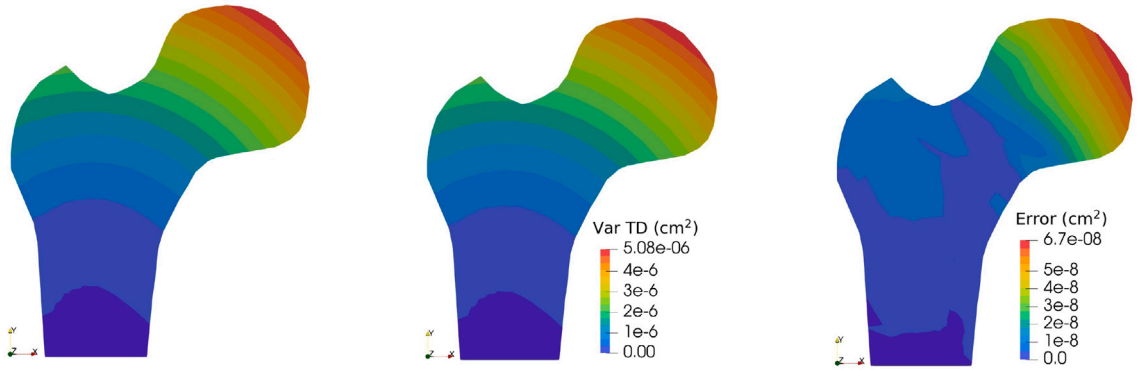


**Fig. 9.** MLMC (left) and MC (middle) variance estimate of total displacement (with identical scale), along with absolute error on the right.

higher number of samples, making it more computationally expensive than the mean estimate. The complete normalization of sampling error reveals non-asymptotic behaviour in the MC cost between mean and variance, while MLMC costs exhibit an asymptotic relationship. Additionally, the difference in accuracy of mean and variance estimates between the MLMC and MC methods is found to be very small.

## CRediT authorship contribution statement

**Sharana Kumar Shivanand:** Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Project administration, Methodology, Investigation, Conceptualization. **Bojana Rosić:** Writing – review & editing, Writing – original draft, Supervision, Methodology, Funding acquisition, Conceptualization.

## Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Sharana Kumar Shivanand reports that financial support was provided by the German Research Foundation. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgment

The authors gratefully acknowledge the financial support of the German Research Foundation (DFG) within the DFG Priority Program SPP 1748 “Reliable Simulation Techniques in Solid Mechanics”.

## Appendix A. h-statistics: unbiased estimation of central moments

The h-statistics, denoted by $\widehat{\mathrm{h}}_p^{\mathrm{MC}}$, serves as an unbiased estimator of the central population moments $\mu_p(u_h)$ in Eq. (5) [14,51]. The ‘MC’ superscript of $\widehat{\mathrm{h}}_p$ signifies that these statistics are obtained through random Monte Carlo-based sampling. For example, the first two h-statistics can be defined as

$$\begin{aligned}\mu_1 &\approx \widehat{\mathrm{h}}_1^{\mathrm{MC}} = 0,\\ \mu_2 &\approx \widehat{\mathrm{h}}_2^{\mathrm{MC}} = \frac{N s_2 - s_1^2}{N(N-1)},\end{aligned} \tag{72}$$

in which $s_1$ and $s_2$ are the power sums, given as

$$s_a(u_h) := \sum_{k=1}^{N} u_h(x,\omega_k)^a, \tag{73}$$

for $a \in \mathbb{Z}_{\geq 0}$. Using the power series representation of h-statistics, the authors in [51] developed a *Mathematica*-based package called *mathStatica*, which efficiently generates h-statistics for any value of $p$.

What distinguishes h-statistics from other unbiased estimators are its properties, such as [25]:

1. **Unbiasedness**: The expectation of the h-statistic equals the corresponding population central moment, i.e., $\mathbb{E}(\widehat{\mathrm{h}}_p^{\mathrm{MC}}) = \mu_p$.
2. **Symmetry**: Among all unbiased estimators of $\mu_p$, $\widehat{\mathrm{h}}_p^{\mathrm{MC}}$ is the only one that exhibits symmetry.[2]
3. **Minimum variance**: Out of all unbiased estimators of $\mu_p$, $\widehat{\mathrm{h}}_p^{\mathrm{MC}}$ stands out for its minimal variance, denoted as $\mathbb{V}\mathrm{ar}(\widehat{\mathrm{h}}_p^{\mathrm{MC}})$.

[2] A function or an estimate is considered symmetric when it is not influenced by the order in which observations are considered.

## Appendix B. Unbiased estimation of $\mu_4$ and $\mu_2^2$

The unbiased approximation of the fourth central moment $\mu_4(u_h)$ is computed by the fourth h-statistic [14]:

$$\begin{aligned}\mu_4(u_h) \approx \mathrm{h}_4^{\mathrm{MC}}(u_h) := \frac{1}{N(N-1)(N-2)(N-3)}\Big( &-3s_1^4 + 6N s_1^2 s_2\\ &+ (9-6\,N)s_2^2 + (-4\,N^2 + 8\,N - 12)s_1 s_3\\ &+ (N^3 - 2\,N^2 + 3\,N)s_4\Big),\end{aligned} \tag{74}$$

whereas the unbiased estimator of $\mu_2(u_h)^2$ is the polyache $\mathrm{h}_{\{2,2\}}^{\mathrm{MC}}$ according to [51,66], given as

$$\begin{aligned}\mu_2(u_h)^2 \approx \mathrm{h}_{\{2,2\}}^{\mathrm{MC}}(u_h) := \frac{1}{N(N-1)(N-2)(N-3)}\Big( &s_1^4 - 2N s_1^2 s_2\\ &+ (N^2 - 3\,N + 3)s_2^2 + (4\,N - 4)s_1 s_3 + (-N^2 + N)s_4\Big).\end{aligned} \tag{75}$$

It is to be noted that the estimator of $\mu_2^2$ by the square of the second h-statistic, i.e., $(\widehat{\mathrm{h}}_2^{\mathrm{MC}})^2$, leads to a biased estimate, whereas only the polyache $\mathrm{h}_{\{2,2\}}^{\mathrm{MC}}$, which is unbiased, remains the preferred way.

## Appendix C. Variance of normalizing factor $\widehat{\lambda}_v := \widehat{\mathbb{V}}_2^{\mathrm{MC}}(u_h)$

One may define the variance of the normalizing factor $\widehat{\lambda}_v := \widehat{\mathbb{V}}_2^{\mathrm{MC}}(u_h)$ using the statistical package *mathStatica* [51] as

$$\begin{aligned}\mathbb{V}\mathrm{ar}(\widehat{\mathbb{V}}_2^{\mathrm{MC}}(u_h)) = &\frac{72\mu_2^4\left(N^2 - 6\,N + 12\right)}{(N-3)(N-2)(N-1)N} + \frac{16\mu_3^2\mu_2\left(N^2 - 4\,N + 13\right)}{(N-2)(N-1)N}\\ &- \frac{24\mu_4\mu_2^2(4\,N-11)}{(N-2)(N-1)N} + \frac{16\mu_6\mu_2}{(N-1)N} + \frac{\mu_8}{N}\\ &- \frac{8\mu_3\mu_5}{N} - \frac{\mu_4^2(N-17)}{(N-1)N}.\end{aligned} \tag{76}$$

Here, the $p$-th central moment is denoted by $\mu_p \equiv \mu_p(u_h)$. For brevity, the previous expression is rewritten as

$$\mathbb{V}\mathrm{ar}(\widehat{\mathbb{V}}_2^{\mathrm{MC}}(u_h)) = \frac{\mathcal{V}_2}{N}, \tag{77}$$

where,

$$\begin{aligned}\mathcal{V}_2 = &\frac{72\mu_2^4\left(N^2 - 6\,N + 12\right)}{(N-3)(N-2)(N-1)} + \frac{16\mu_3^2\mu_2\left(N^2 - 4\,N + 13\right)}{(N-2)(N-1)}\\ &- \frac{24\mu_4\mu_2^2(4\,N-11)}{(N-2)(N-1)} + \frac{16\mu_6\mu_2}{N-1} + \mu_8 - 8\mu_3\mu_5 - \frac{\mu_4^2(N-17)}{N-1}.\end{aligned} \tag{78}$$

Eq. (77) indicates that the statistical error of the estimator $\widehat{\mathbb{V}}_2^{\mathrm{MC}}(u_h)$ converges at a rate of $\mathcal{O}(N^{-1})$.

## Appendix D. Unbiased estimation of $\mathbb{V}_{l,2}$

Following the notion of power sum $s_a$ defined in Appendix A, here we introduce the bivariate power sum [35,51]

$$s_{a,b} := \sum_{i=1}^{N} X_{h_l}^{+}(\omega_i)^a X_{h_l}^{-}(\omega_i)^b, \tag{79}$$

where $X_{h_l}^{+}(\omega_i)_{i=1,\ldots,N_l} := X_{h_l,N_l}^{+} = u_{h_l,N_l} + u_{h_{l-1},N_l}$ and $X_{h_l}^{-}(\omega_i)_{i=1,\ldots,N_l} := X_{h_l,N_l}^{-} = u_{h_l,N_l} - u_{h_{l-1},N_l}$. Subsequently, the unbiased MC estimation of

the quantity $\mathbb{V}_{l,2}$ reads:

$$\begin{aligned}\widehat{\mathbb{V}}^{\mathrm{MC}}_{l,2} = \frac{1}{(N_l-3)(N_l-2)(N_l-1)^2N_l}\Big(&N_l\big((-N_l^2+N_l+2)(s_{1,1})^2\\ &+(N_l-1)^2(N_l s_{2,2}-2s_{1,0}s_{1,2})+(N_l-1)s_{0,2}((s_{1,0})^2-s_{2,0})\big)\\ &+(s_{0,1})^2\big((6-4N_l)(s_{1,0})^2+(N_l-1)N_l s_{2,0}\big)\\ &-2N_l s_{0,1}\big((N_l-1)^2 s_{2,1}+(5-3N_l)s_{1,0}s_{1,1}\big)\Big).\end{aligned} \tag{80}$$

## Data availability

Data will be made available on request.

## References


[1] Ashman RB, Cowin SC, Van Buskirk WC, Rice JC. A continuous wave technique for the measurement of the elastic properties of cortical bone. J Biomech January. 1984;17(5):349–61. https://doi.org/10.1016/0021-9290(84)90029-0

[2] Barth A, Schwab C, Zollinger N. Multi-level monte carlo finite element method for elliptic PDES with stochastic coefficients. Numer Math 2011;119(1):123–61. https://doi.org/10.1007/s00211-011-0377-0

[3] Betz W, Papaioannou I, Straub D. Numerical methods for the discretization of random fields by means of the karhunen–loève expansion. Comput Methods Appl Mech Eng April. 2014;271:109–29. https://doi.org/10.1016/j.cma.2013.12.010

[4] Bierig C, Chernov A. Convergence analysis of multilevel monte carlo variance estimators and application for random obstacle problems. Numer Math August. 2015;130(4):579–613. https://doi.org/10.1007/s00211-014-0676-3

[5] Bierig C, Chernov A. Estimation of arbitrary order central statistical moments by the multilevel monte carlo method. Stochastics Partial Differ Equ Anal Comput March. 2016;4(1):3–40. https://doi.org/10.1007/s40072-015-0063-9

[6] Bóna A, Bucataru I, Slawinski MA. Coordinate-free characterization of the symmetry classes of elasticity tensors. J Elast June 2007;87(2–3):109–32. https://doi.org/10.1007/s10659-007-9099-z

[7] Charrier J, Scheichl R, Teckentrup AL. Finite element error analysis of elliptic PDES with random coefficients and its application to multilevel monte carlo methods. SIAM J Numer Anal 2013;51(1):322–52. https://doi.org/10.1137/110853054

[8] Cho E, Cho M. Variance of sample variance with replacement. Int J Pure Appl Math January 2009;52:43–7.

[9] Cliffe KA, Giles MB, Scheichl R, Teckentrup AL. Multilevel monte carlo methods and applications to elliptic PDES with random coefficients. Comput Vis Sci 2011;14(1):3–15. https://doi.org/10.1007/s00791-011-0160-x

[10] Collier N, Haji-Ali A-L, Nobile F, von Schwerin E, Tempone R. A continuation multilevel monte carlo algorithm. BIT Numerical Mathematics June 2015;55(2):399–432. https://doi.org/10.1007/s10543-014-0511-3

[11] Cowin SC. Continuum mechanics of anisotropic materials. New York, NY: Springer New York; 2013. https://doi.org/10.1007/978-1-4614-5025-2

[12] Crow EL. Lognormal distributions: theory and applications. 1st ed. Routledge; 1988. https://doi.org/10.1201/9780203748664

[13] DeCarlo LT. On the meaning and use of kurtosis. Psychol Methods 1997;2(3):292–307. https://doi.org/10.1037/1082-989X.2.3.292

[14] Dwyer PS. Moments of any rational integral isobaric sample moment function. Ann Math Stat 1937;8(1):21–65. Publisher: Institute of Mathematical Statistics.

[15] Fischer H. A history of the central limit theorem: from classical to modern probability theory. In: Sources and studies in the history of Mathematics and physical sciences. 1 ed. New York, NY: Springer; 2011. https://doi.org/10.1007/978-0-387-87857-7

[16] Fishman G. Monte carlo: concepts, algorithms, and applications. In: Springer series in operations research and financial engineering. New York: Springer-Verlag; 1996. https://doi.org/10.1007/978-1-4757-2553-7

[17] Gaillac R, Pullumbi P, Coudert F-X. ELATE: an open-source online application for analysis and visualization of elastic tensors. J Phys Condens Matter May 2016. Publisher: IOP Publishing. doi: 28(27):275201. https://doi.org/10.1088/0953-8984/28/27/275201

[18] Geraldes D, Phillips A. A comparative study of orthotropic and isotropic bone adaptation in the femur. Int J Numer Methods Biomed Eng 2014;30(9):873–89. https://doi.org/10.1002/cnm.2633

[19] Ghanem RG, Spanos PD. Stochastic finite elements: a spectral approach. Berlin, Heidelberg: Springer-Verlag; 1991. https://doi.org/10.1007/978-1-4612-3094-6

[20] Giles MB. Multilevel monte carlo path simulation. Operations Research 2008;56(3):607–17. https://doi.org/10.1287/opre.1070.0496

[21] Giles MB. Multilevel monte carlo methods. Acta Numerica 2015;24:259–328. https://doi.org/10.1017/S096249291500001X

[22] Graham C, Talay D. Stochastic simulation and monte carlo methods: mathematical foundations of stochastic simulation. Stochastic Modelling and Applied Probability. Berlin Heidelberg: Springer-Verlag; 2013. https://doi.org/10.1007/978-3-642-39363-1

[23] Guilleminot J, Soize C. A stochastic model for elasticity tensors with uncertain material symmetries. Int J Solids Struct 2010;47(22):3121–30. https://doi.org/10.1016/j.ijsolstr.2010.07.013

[24] Haji-Ali A-L, Nobile F, Von Schwerin E, Tempone R. Optimization of mesh hierarchies in multilevel monte carlo samplers. Stochastics Partial Differ Equ Anal Comput March. 2016;4(1):76–112. https://doi.org/10.1007/s40072-015-0049-7

[25] Halmos PR. The theory of unbiased estimation. Ann Math Stat 1946. Publisher: Institute of Mathematical Statistics. doi: 17(1):34–43. https://doi.org/10.1214/aoms/1177731020

[26] Heinrich S. Multilevel monte carlo methods. In: Large-scale scientific computing, lecture notes in computer science. Springer; 2001. p. 58–67. https://doi.org/10.1007/3-540-45346-6_5

[27] Hammersley JM, Handscomb DC. Monte carlo methods. In: Monographs on statistics and applied probability. Dordrecht: Springer; 1964. https://doi.org/10.1007/978-94-009-5819-7

[28] Kamiński M. Generalized perturbation-based stochastic finite element method in elastostatics. Computers Struct 2007;85(10):586–94. https://doi.org/10.1016/j.compstruc.2006.08.077

[29] Karhunen K. Über lineare methoden in DER wahrscheinlichkeitsrechnung. Ann Acad Sci Fenn 1947;37:3–79.

[30] Keese A. Numerical solution of systems with stochastic uncertainties: a general purpose framework for stochastic finite elements. [Ph.D. thesis], Technische Universität Braunschweig; 2004. https://doi.org/10.24355/dbbs.084-200511080100-436

[31] Keller TS. Predicting the compressive mechanical behavior of bone. J Biomech 1994;27(9):1159–68. https://doi.org/10.1016/0021-9290(94)90056-6

[32] Kenney JF, Keeping ES. Mathematics of statistics. 3rd ed. New York: Van Nostrand; 1954.

[33] Kleiber M, Hien TD. The stochastic finite element method: basic perturbation technique and computer implementation. Chichester: John Wiley & Sons; 1992.

[34] Kosambi DD. Statistics in function space. J Indian Math Soc 1943;7:76–88.

[35] Krumscheid S, Nobile F, Pisaroni M. Quantifying uncertain system outputs via the multilevel monte carlo method — Part I: Central Moment Estimation. J Comput Phys August. 2020;414:109466. https://doi.org/10.1016/j.jcp.2020.109466

[36] Liu WK, Belytschko T, Mani A. Probabilistic finite elements for nonlinear structural dynamics. Comput Methods Appl Mech Eng 1986;56(1):61–81. https://doi.org/10.1016/0045-7825(86)90136-2

[37] Loève M. Fonctions aleatoire du second ordre, supplement to p. Levy. In: Processus stochastic ET mouvement brownien. Gauthier-Villars; 1948, Paris.

[38] Lord GJ, Powell CE, Shardlow T. An introduction to computational stochastic PDES. In: Cambridge texts in applied Mathematics. Cambridge University Press; 2014. https://doi.org/10.1017/CBO9781139017329

[39] Malyarenko A, Ostoja-Starzewski M. Tensor-valued random fields for continuum Physics. Cambridge University Press; 2018. https://doi.org/10.1017/9781108555401

[40] Marsaglia G, Tsang WW. A fast, easily implemented method for sampling from decreasing or symmetric unimodal density functions. SIAM J Sci Stat Comput 1984;5(2):349–59. https://doi.org/10.1137/0905026

[41] Marsaglia G, Tsang WW. The ziggurat method for generating random variables. J Stat Softw October. 2000;5:1–7. https://doi.org/10.18637/jss.v005.i08

[42] Marsden JE, Hughes TJR. Mathematical foundations of elasticity. Englewood Cliffs, NJ: Prentice-Hall, Inc.; 1983.

[43] Matthies HG, Brenner CE, Bucher CG, Soares CG. Uncertainties in probabilistic numerical analysis of structures and solids-stochastic finite elements. Structural Safety 1997;19(3):283–336. https://doi.org/10.1016/S0167-4730(97)00013-1

[44] Matthies HG, Bucher C. Finite elements for stochastic media problems. Comput Methods Appl Mech Eng 1999;168(1):3–17. https://doi.org/10.1016/S0045-7825(98)00100-5

[45] Matthies HG, Keese A. Galerkin methods for linear and nonlinear elliptic stochastic partial differential equations. Comput Methods Appl Mech Eng April. 2005;194(12):1295–331. https://doi.org/10.1016/j.cma.2004.05.027

[46] Menhorn F, Geraci G, Seidl DT, Marzouk YM, Eldred MS, Bungartz H-J. Multilevel monte carlo estimators for derivative-free optimization under uncertainty. Int J Uncertain Quantif 2024;14(3). https://doi.org/10.1615/Int.J.UncertaintyQuantification.2023048049

[47] Metropolis N, Ulam S. The monte carlo method. J Am Stat Assoc 1949;44(247):335–41.

[48] Mishra S, Schwab C, Šukys J. Multi-level monte carlo finite volume methods for nonlinear systems of conservation laws in multi-dimensions. J Comput Phys April. 2012;231(8):3365–88. https://doi.org/10.1016/j.jcp.2012.01.011

[49] Ostoja-Starzewski M. Microstructural randomness and scaling in mechanics of materials. New York: Chapman and Hall/CRC. August. 2007. https://doi.org/10.1201/9781420010275

[50] Papadrakakis M, Papadopoulos V. Robust and efficient methods for stochastic finite element analysis using monte carlo simulation. Comput Methods Appl Mech Eng 1996;134(3):325–40. https://doi.org/10.1016/0045-7825(95)00978-7

[51] Rose C, Smith MD. Mathematical statistics with *Mathematica*, vol. 481. New York, NY: Springer; 2002.

[52] Rosić B. Variational formulations and functional approximation algorithms in stochastic plasticity of materials. [Ph.D. thesis], Technische Universität Braunschweig; 2012. https://doi.org/10.24355/dbbs.084-201307240922-0

[53] Sansalone V, Gagliardi D, Desceliers C, Bousson V, Laredo J, Peyrin F, Haïat G, Naili S. Stochastic multiscale modelling of cortical bone elasticity based on high-resolution imaging. Biomech Model Mechanobiol February 2016;15(1):111–31. https://doi.org/10.1007/s10237-015-0695-8

[54] Schuëller GI. Computational stochastic mechanics – recent advances. Comput Struct 2001;79(22):2225–34. https://doi.org/10.1016/S0045-7949(01)00078-5

[55] Shinozuka M, Deodatis G. Response variability of stochastic finite element systems. J Eng Mech March. 1988. Publisher: American Society of Civil Engineers. doi: 114(3):499–519. https://doi.org/10.1061/(ASCE)0733-9399(1988)114:3(499)
[56] Shinozuka M, Yamazaki F. Stochastic finite element analysis: an introduction. In: Ariaratnam ST, Schuëller GI, Elishakoff I, editors. Stochastic structural dynamics. Elsevier Applied Science Publishers Ltd; 1995.
[57] Shivanand SK. Covariance estimation using h-statistics in monte carlo and multilevel monte carlo methods. Int J Uncertain Quantif 2025;15(2):43–64. https://doi.org/10.1615/Int.J.UncertaintyQuantification.2024051528
[58] Shivanand SK, Rosić B, Matthies HG. Stochastic modelling of symmetric positive definite material tensors. J Comput Phys May 2024;505:112883. https://doi.org/10.1016/j.jcp.2024.112883
[59] Shivanand SK, Rosić B, Matthies HG. Stochastic modelling of elasticity tensor fields. Math Mech Solids 2025. https://doi.org/10.1177/10812865251348030
[60] Sobczyk K, Kirkner DJ. Stochastic modeling of microstructures. In: Modeling and simulation in science, engineering and technology. Boston: Birkhäuser; 2001. https://doi.org/10.1007/978-1-4612-0121-2
[61] Soize C. Non-gaussian positive-definite matrix-valued random fields for elliptic stochastic partial differential operators. Comput Methods Appl Mech Eng 2006;195(1):26–64. https://doi.org/10.1016/j.cma.2004.12.014
[62] Soize C. Tensor-valued random fields for meso-scale stochastic model of anisotropic elastic microstructure and probabilistic analysis of representative volume element size. Probabilist Eng Mech 2008;23(2):307–23. https://doi.org/10.1016/j.probengmech.2007.12.019
[63] Spanos PD, Ghanem R. Stochastic finite element expansion for random media. J Eng Mech 1989;115(5):1035–53. https://doi.org/10.1061/(ASCE)0733-9399(1989)115:5(1035)
[64] Stefanou G. The stochastic finite element method: past, present and future. Comput Methods Appl Mech Eng 2009;198(9):1031–51. https://doi.org/10.1016/j.cma.2008.11.007
[65] Teckentrup AL, Scheichl R, Giles MB, Ullmann E. Further analysis of multilevel monte carlo methods for elliptic PDES with random coefficients. Numer Math November. 2013;125(3):569–600. https://doi.org/10.1007/s00211-013-0546-4
[66] Tracy DS, Gupta BC. Generalized *h*-statistics and other symmetric functions. Ann Stat 1974. Publisher: Institute of Mathematical Statistics. doi: 2(4). https://doi.org/10.1214/aos/1176342774
[67] Xiu D. Numerical methods for stochastic computations: a spectral method approach. Princeton University Press; 2010.
[68] Yamazaki F, Shinozuka M, Dasgupta G. Neumann expansion for stochastic finite element analysis. J Eng Mech August. 1988. Publisher: American Society of Civil Engineers. doi: 114(8):1335–54. https://doi.org/10.1061/(ASCE)0733-9399(1988)114:8(1335)
[69] Yosibash Z, Padan R, Joskowicz L, Milgrom C. A CT-based high-order finite element analysis of the human proximal femur compared to in-vitro experiments. J Biomech Eng June 2007;129(3):297–309. https://doi.org/10.1115/1.2720906
[70] Zienkiewicz OC, Taylor RL. The finite element method: its basis and fundamentals. In: The finite element method. Butterworth-Heinemann; 2013.