2152–5080/17/$35.00

EMBEDDED NONLOCAL OPERATOR REGRESSION (ENOR): QUANTIFYING MODEL ERROR IN LEARNING NONLOCAL OPERATORS

Yiming Fan,[1] Habib N. Najm,[2] Yue Yu,[1,*] Stewart Silling,[3] & Marta D'Elia[4]

[1]*Department of Mathematics, Lehigh University, Bethlehem, PA 18015, USA.*

[2]*Sandia National Laboratories, Livermore, CA, USA.*

[3]*Center for Computing Research, Sandia National Laboratories, Albuquerque, NM, USA.*

[4]*Stanford University, Stanford, CA, USA.*

*Address all correspondence to: Yue Yu, Corresponding author. Department of Mathematics, Lehigh University, Bethlehem, PA 18015, USA., E-mail: yuy214@lehigh.edu

Nonlocal, integral operators have become an efficient surrogate for bottom-up homogenization, due to their ability to represent long-range dependence and multiscale effects. However, the nonlocal homogenized model has unavoidable discrepancy from the microscale model. Such errors accumulate and propagate in long-term simulations, making the resultant prediction unreliable. To develop a robust and reliable bottom-up homogenization framework, we propose a new framework, which we coin Embedded Nonlocal Operator Regression (ENOR), to learn a nonlocal homogenized surrogate model and its structural model error. This framework provides discrepancy-adaptive uncertainty quantification for homogenized material response predictions in long-term simulations. The method is built on Nonlocal Operator Regression (NOR), an optimization-based nonlocal kernel learning approach, together with an embedded model error term in the trainable kernel. Then, Bayesian inference is employed to infer the model error term parameters together with the kernel parameters. To make the problem computationally feasible, we use a multilevel delayed acceptance Markov chain Monte Carlo (MLDA-MCMC) method, enabling efficient Bayesian model calibration and model error estimation. We apply this technique to predict long-term wave propagation in a heterogeneous one-dimensional bar, and compare its performance with additive noise models. Owing to its ability to capture model error, the learned ENOR achieves improved estimation of posterior predictive uncertainty.

KEY WORDS: *Model Error, Nonlocal Model, Uncertainty Quantification, Bayesian Inference*

Contents

1. INTRODUCTION

In many real-world applications, the analyzed physical system is a complex multi-scale process, which often starts from a sequence of mechanical cues at the microscale, and results in the propagation of properties at the macroscle. However, microscale simulations are often infeasible due to high computational costs. To mitigate this challenge, upscaled models are often needed, which can be employed in large scale long-term simulations to inform decision making. For this purpose, various multiscale approaches and homogenized surrogate models were developed [1–11]. Examples include partial differential equation (PDE)-based approaches with effective coefficients [12,13], particle-based methods that replace clusters of particles with larger particles [14–16], nonlocal models that capture long-range microscale effects via integral operators [17–27], and several others. Among these approaches, nonlocal models [28–34] stand out as relatively recent yet powerful tools [17–20,27,35–38] as their integral nature naturally embeds length and time-scales in their definition [28] allowing them to capture macroscale effects induced by a heterogeneous microscale behavior.

However, despite being appropriate for homogenization tasks, nonlocal models present several challenges. As an example, in heterogeneous materials modeling, the material's response depends on mechanical and microstructural properties, and hence requires a material-specific homogenized nonlocal model. In practice, a nonlocal model is determined by the definition of its "kernel" whose functional form and associated parameters are not known *a priori*. Thus, to accurately capture the effects of micro-scale heterogeneities at the macroscale and provide reliable and trustworthy predictions, it is necessary to estimate appropriate nonlocal kernels. The paper [39] introduces the nonlocal operator regression (NOR) approach, a data-driven technique for the identification of the nonlocal kernel that best describes a system at the macroscale. We refer the reader to [22,39–45] for several examples of machine learning-based design of homogenized nonlocal operators, and to [25,46] for the rigorous analysis of its learning theory.

While NOR provides accurate and reliable deterministic nonlocal homogenization models from data, it introduces unavoidable modeling errors due to the discrepancy between the "surrogate" nonlocal model and the ground truth system. In long-term simulations, such a modeling error may propagate and accumulate over time. Hence, it is desired to characterize the discrepancy between the nonlocal model and the true governing equations to estimate the predictions' uncertainty. In our prior work [47], we introduced the Bayesian nonlocal operator regression (BNOR) approach, which combines Bayesian inference [48–59] and NOR to estimate model uncertainty under the assumption of additive independent Gaussian noise in the data. Specifically, we modeled the discrepancy between high-fidelity data and the NOR model as additive independent identically distributed (*iid*) noise, and tested the method in the context of stress wave propagation in a randomly heterogeneous material [47]. However, the additive *iid* noise assumption is more suited for

representing measurement noise rather than a modeling error; this happens because the model discrepancy is expected to exhibit a significant degree of spatial correlation depending on the degree of smoothness in the solution and data. Ignoring this correlation may lead to over-confidence and low posterior uncertainty in the inferred parameters, with consequent over confidence in posterior predictions of the quantities of interest (QoIs) [60].

To better represent the discrepancies due to model error, state-of-the-art additive model error techniques augment the predictive model's output with a Gaussian process (GP)[54,61]. The additive GP provides extensive flexibility in correcting model outputs and fitting available data with a parameterized correlation structure, resulting in meaningful predicted uncertainty. However, adding a GP to model outputs in the context of physical systems may compromise adhering to physical constraints and governing equations; in fact, while a physical model prediction $y = f(x)$ is expected to satisfy relevant physical laws, a GP-augmented prediction $y = f(x) + \epsilon(x)$ may not in principle [62].

To address the shortcomings of additive GP constructions in physical models, embedded model error constructions were introduced in [62,63]; here, specific model components are augmented with statistical terms. Specifically, random variables are embedded in the parameterization of certain model elements based on approximations or modeling assumptions. As a result, this approach enables Bayesian parameter estimation that accounts for model error and identifies modeling assumptions that dominate the discrepancies between the fitted model and the data. To our knowledge, embedded model error applications have relied on random variable embeddings and not GP embeddings.

In this work, we introduce a Bayesian calibration technique based on GP embeddings applied in the context of nonlocal homogenization. In particular, we augment the nonlocal kernel with a GP to represent model error in the kernel. Then, we use Bayesian inference to learn the posterior probability distributions of parameters of the nonlocal constitutive law (the nonlocal kernel function) and the GP simultaneously. To make the solution of the resulting Bayesian inference problem feasible, we employ the Multilevel Delayed Acceptance (MLDA) MCMC [64,65] method, which exploits a hierarchy of models with increasing complexity and cost. We illustrate this method using a one-dimensional nonlocal wave equation that describes the propagation of stress waves through an elastic bar with a heterogeneous microstructure.

We summarize our contributions in the following list.

- We propose the Embedded Nonlocal Operator Regression (ENOR) approach, which consist of augmenting a nonlocal homogenized model with modeling error with the purpose of capturing discrepancy of the predictions with respect to high fidelity data. In particular, we embed a GP in the nonlocal kernel to characterize the spatial heterogeneity in the microscale.

- The model is inferred from high-fidelity microscale data using MCMC, which learns the parameters of the nonlocal kernel and the correction term jointly. To alleviate the cost associated with costly likelihood estimations in MCMC, we use a multi-level delayed acceptance MCMC formulation.

- To illustrate the efficacy of our method, we consider the stress wave propagation problem in a heterogeneous one-dimensional bar. A data-driven nonlocal homogenized surrogate is obtained, together with a GP representing the kernel correction. Compared to the previous work which learns an additive noise [47], the embedded model error correction formulation successfully captures the discrepancy and provides an improved estimation of the posterior predictive uncertainty.

Paper outline. In Section 2, we review the nonlocal operator regression approach and the general embedded model error representation formulation. Then, in Section 3 we propose the ENOR approach, together with a MLDA-MCMC formulation which uses a low-fidelity model to accelerate the expensive likelihood estimation. In Section 4, we examine the posterior distribution and the prediction given by the algorithm. Section 5 concludes the paper with a summary of our contributions and a discussion of potential follow-up work.

2. BACKGROUND AND RELATED MATHEMATICAL FORMULATION

Although the proposed method is generalizable to other homogenization problems, for the purpose of demonstration, in this paper we consider the numerical solution of a wave propagation problem in the spatial domain $\bar{\Omega} \subset \mathbb{R}^d$ and time domain $[0, T]$, where d is the physical dimension. Given S observations of forcing terms $f^s(x,t)$, we define the corresponding computed high-fidelity displacement fields $\mathcal{D} := \{u^s_{DNS}\}^S_{s=1}$ as the ground-truth dataset. Here we assume that both f^s and u^s_{DNS} are provided at time instants $t^n \in [0,T]$ and grid points $x_i \in \bar{\Omega}$. Without loss of generality, we assume the time and space points to be uniformly distribute, i.e. in one dimension, the spatial grid size Δx and time step size Δt are constant. We denote the collection of all grid points as $\chi = \{x_i\}^L_{i=1}$. The purpose of this work is to develop a Bayesian inference framework, which learns the nonlocal model from $\mathcal{D}$ together with its structural error.

Throughout this paper, for any vector $\mathbf{v} = [v_1, \cdots, v_q] \in \mathbb{R}^q$, we use $||\mathbf{v}||_{l^2} := \sqrt{\sum_{i=1}^q v_i^2}$ to denote its l^2 norm. For a function $u(x,t)$ with $(x,t) \in \bar{\Omega} \times [0,T]$, its discrete l^2 norm is defined as

$$||u||_{l^2(\bar{\Omega}\times[0,T])} := \sqrt{\Delta t \Delta x \sum_{n=0}^{T/\Delta t} \sum_{x_i \in \chi} (u(x_i, t^n))^2},$$

which can be interpreted as a numerical approximation of the $L^2(\bar{\Omega} \times [0,T])$ norm of u.

2.1 Nonlocal Operator Regression

In this work, we consider the nonlocal operator regression (NOR) approach first proposed in [40] and later extended in [22,25,39,42,46]. NOR aims to find the best homogenized nonlocal "surrogate" model from experimental and/or high-fidelity simulation data pairs. Herein, we focus on the latter scenario. Given a forcing term $f(x,t)$, $(x,t) \in \bar{\Omega} \times [0,T]$ and appropriate boundary and initial conditions, we denote the high-fidelity (HF) model as

$$\frac{\partial^2 u_{HF}}{\partial t^2}(x,t) - \mathcal{L}_{HF}[u_{HF}](x,t) = f(x,t). \tag{1}$$

$\mathcal{L}_{HF}$ is the HF operator and $u_{HF}(x,t)$ is the solution of fine-scale simulations. NOR assumes that nonlocal models are accurate homogenized surrogates of the HF model; a general nonlocal model reads as follows:

$$\frac{\partial^2 u_{NL}}{\partial t^2}(x,t) - \mathcal{L}_{NL}[u_{NL}](x,t) = f(x,t), \tag{2}$$

where the nonlocal operator $\mathcal{L}_{NL}$ is an integral operator of the form

$$\mathcal{L}_{NL} := \mathcal{L}_{K_{\mathbf{C}}}[u](x,t) = \int_{\bar{\Omega} \cap B_\delta(x)} K_{\mathbf{C}}(x,y)(u(y,t) - u(x,t))dy. \tag{3}$$

Here, $B_\delta(x) := \{y \in \mathbb{R}^d,\ |x-y| < \delta\}$ denotes the interaction neighborhood of the material point x. We further denote $\Omega_\delta := \{x \in \bar{\Omega} \,|\, \text{dist}(x, \mathbb{R}^d \backslash \bar{\Omega}) < \delta\}$ is the nonlocal boundary region, $\Omega := \bar{\Omega} \backslash \Omega_\delta$ as the "interior" region inside the domain. $\mathbf{C} := \{C_m\}_{m=0}^M$ is the set of parameters that uniquely determines the kernel $K_{\mathbf{C}}$. We claim that for the same forcing terms $f(x,t)$, the corresponding solution, $u_{NL}(x,t)$, $(x,t) \in \bar{\Omega} \times [0,T]$, of a nonlocal model provides an approximation of the ground-truth data, i.e., $u_{NL}(x,t) \approx u_{DNS}(x,t)$. Here, a radial kernel, $K_{\mathbf{C}}(x,y) := K_{\mathbf{C}}(|y-x|)$, is employed. This widely adopted setting guarantees the symmetry of the integrand in (3) with respect to x and y and induces the conservation of linear momentum and Galilean invariance [38,42,66,67]. In fact, such property allows us to write:

$$K_{\mathbf{C}}(x,y)(u(y,t) - u(x,t)) = \frac{1}{2}\left[K_{\mathbf{C}}(|x-y|)(u(y,t) - u(x,t)) - K_{\mathbf{C}}(|y-x|)(u(x,t) - u(y,t))\right]. \tag{4}$$

As done in previous works[47], we represent the kernel using a linear combination of Bernstein-polynomials, i.e.,

$$K_{\mathbf{C}}\left(\frac{|z|}{\delta}\right)=\sum_{m=0}^{M}\frac{C_m}{\delta^{d+2}}B_{m,M}\left(\left|\frac{z}{\delta}\right|\right),\quad B_{m,M}(z)=\binom{M}{m}z^m(1-z)^{M-m}\quad \text{for } 0\le z\le 1. \tag{5}$$

Note that this construction makes $K_{\mathbf{C}}$ continuous, radial and compactly supported on the ball of radius δ centered at x, also guarantees that (2) is well-posed [68].

To find the nonlocal model that best describes the high-fidelity data, we aim to find the kernel parameters $\mathbf{C}$ such that for given loadings $f^s(x,t)$, the corresponding nonlocal solutions $u^s_{NL,\mathbf{C}}(x,t)$ in (2) are as close as possible to the HF solution $u^s_{DNS}(x,t)$. To numerically solve (2), we use a mesh-free discretization of the nonlocal operator and a central difference scheme in time, i.e.

$$\begin{aligned}(u^s_{NL,\mathbf{C}})_i^{n+1} &:= 2(u^s_{NL,\mathbf{C}})_i^n-(u^s_{NL,\mathbf{C}})_i^{n-1}+\Delta t^2 f^s(x_i,t^n)+\Delta t^2\left(\mathcal{L}_{K_{\mathbf{C}},h}[u^s_{NL,\mathbf{C}}]\right)_i^n\\ &=2(u^s_{NL,\mathbf{C}})_i^n-(u^s_{NL,\mathbf{C}})_i^{n-1}+\Delta t^2 f^s(x_i,t^n)\\ &+\Delta t^2\Delta x\sum_{x_j\in B_\delta(x_i)\cap\chi}K_{\mathbf{C}}(|x_j-x_i|)((u^s_{NL,\mathbf{C}})_j^n-(u^s_{NL,\mathbf{C}})_i^n),\end{aligned} \tag{6}$$

where $(\tilde{u}^s_{NL,\mathbf{C}})_i^n$ denotes the nonlocal solution at (x_i,t^n), and $\mathcal{L}_{K_{\mathbf{C}},h}$ is an approximation of $\mathcal{L}_{K_{\mathbf{C}}}$ by the Riemann sum with uniform grid spacing Δx. The optimal parameters $\mathbf{C}^*$ can be obtained by solving the following optimization problem

$$\mathbf{C}^*=\underset{\mathbf{C}}{\operatorname{argmin}}\sum_{s=1}^{S}\frac{||u^s_{NL,\mathbf{C}}-u^s_{DNS}||^2_{l^2(\Omega\times[0,T])}}{||u^s_{DNS}||^2_{l^2(\Omega\times[0,T])}}+\lambda||\mathbf{C}||^2_{l^2}, \tag{7}$$

$$\text{s.t. } K_{\mathbf{C}} \text{ satisfies physics-based constraints.} \tag{8}$$

Here, λ is a regularization parameter added to guarantee the well-conditioning of the optimization problem. The physics-based constraints depend on the nature of the problem, in the case of stress waves, we impose homogenized properties of plane waves propagating at very low frequency [39], i.e.

$$\int_{B_\delta(x)}|y-x|^2K_{\mathbf{C}}(|y-x|)dy=2\rho c_0^2, \tag{9}$$

$$\int_{B_\delta(x)}|y-x|^4K_{\mathbf{C}}(|y-x|)dy=-8\rho c_0^3R. \tag{10}$$

The density ρ, the effective wave speed c_0 and R, the second derivative of the wave group velocity with respect to

the frequency ω evaluated at $\omega = 0$, are obtained the same way as in [47], for both periodic and random materials. Without loss of generality, we demonstrate the idea in 1D wave propagation problems. Numerically, we impose the physics constraints by approximating (9) and (10) via a 1D Riemann sum:

$$\sum_{x_j \in B_\delta(x_i) \cap \chi} |x_i - x_i|^2 K_{\mathbf{C}}(|x_j - x_i|)\Delta x = 2\rho c_0^2, \tag{11}$$

$$\sum_{x_j \in B_\delta(x_i) \cap \chi} |x_i - x_i|^4 K_{\mathbf{C}}(|x_j - x_i|)\Delta x = -8\rho c_0^3 R. \tag{12}$$

By solving the above linear equations, one can explicitly express C_{M-1} and C_M as linear functions of other parameters. Then, (7) becomes an unconstrained optimization problem of $\{C_m\}_{m=0}^{M-2}$. For further details we refer interested readers to [25].

We note that in (6), the numerical approximation $(u^s_{NL,\mathbf{C}})^{n+1}$ is calculated based on the numerical approximation of the last time instance $(u^s_{NL,\mathbf{C}})^n_i$. As such, the numerical error accumulates as n increases, and (7) aims to minimize the accumulated error. Alternatively, one can also choose to minimize the step-wise error [25,46], by replacing the approximated solution at the last time instances, $(u^s_{NL,\mathbf{C}})^n$ and $(u^s_{NL,\mathbf{C}})^{n-1}$, with the corresponding ground-truth data $(u^s_{DNS})^n$ and $(u^s_{DNS})^{n-1}$ in (6):

$$\begin{aligned}(\tilde{u}^s_{NL,\mathbf{C}})^{n+1}_i :=& \, 2(u^s_{DNS})^n_i - (u^s_{DNS})^{n-1}_i + \Delta t^2 f^s(x_i, t^n) + \Delta t^2 \left(\mathcal{L}_{K_{\mathbf{C}},h}[u^s_{DNS}]\right)^n_i \\ =& \, 2(u^s_{DNS})^n_i - (u^s_{DNS})^{n-1}_i + \Delta t^2 f^s(x_i, t^n) + \Delta t^2 \Delta x \sum_{x_j \in B_\delta(x_i) \cap \chi} K_{\mathbf{C}}(|x_j - x_i|)((u^s_{DNS})^n_j - (u^s_{DNS})^n_i).\end{aligned}$$

This formula provides a faster approximated nonlocal solution at the cost of physical stability, which plays a critical role in long-term prediction. Therefore, in this work we aim to minimize the accumulated error by considering (6).

2.2 Embedded Model Error

In this section we introduce the embedded model error procedure developed in [62,63], which relies on a Bayesian inference framework for the estimation of the model error, with the identification of the contribution of different error sources to predictive uncertainty. This is a natural setting for calibrating a low fidelity (LF/here NL) model against a higher fidelity (HF/here DNS) model accounting for uncertainty. Here, we assume that any discrepancy between the two models' predictions is due to model error, and not to other sources such as measurement error. Thus, to model this discrepancy, we do not pursue traditional approaches that introduce independent identically distributed (*iid*) additive

Gaussian noise. Kennedy and O'Hagan [69] introduced the use of an additive Gaussian process (GP) to capture the structure of the discrepancy between two models, where the flexibility of the GP allows one to adequately capture the discrepancy in the predictions induced by the model error.

In this setting, denoting by q_j the data generated from the HF/DNS model $\eta(o_j)$, and by $\zeta(o_j, \Lambda)$ the one generated from the LF/NL model, where Λ is a set of model parameters to be estimated, at operating conditions o_j (e.g. spatial or temporal coordinates), for $j = 1, 2, ..., N$ observations, the additive GP construction can be written as

$$q_j = \zeta(o_j, \Lambda) + \Delta(x_j, \alpha, \xi) \tag{13}$$

where $\Delta(x_j, \alpha, \xi)$ is a GP with parameters α evaluated at the discrete spatial locations x_j, and with *iid* standard normal stochastic degrees of freedom ξ. This representation and the related approach have found extensive use, *e.g.* [70–72], as they allow for avoiding the overconfidence that comes from ignoring model error in model calibration whether against another model or actual observations. However, the additive GP construction presents some challenges in calibrating physical models [62], which motivated embedding model error terms into the LF model. In our present notation, this can be written as

$$q_j \approx h(o_j, \tilde{\alpha}, \xi) := \zeta(o_j, \Lambda, \Delta(x_j, \alpha, \xi)), \qquad j = 1...N \tag{14}$$

where $\tilde{\alpha} = (\Lambda, \alpha)$. We note that in the literature the use of model error embedding has relied on a simplified version of the above, where the error term is the random variable $\Delta(\alpha, \xi)$, *i.e.*, lacking the spatial dependence, rather than a GP [62,63,73–75]. This simplification might be necessary because of lack of data or computational constraints and it comes with loss of flexibility in the model error representation. In this work, we retain the full GP formalism, as in (14). It is also important to point out that one key benefit from model error embedding is that the analyst, knowing where approximations have been made in the computational model at hand, can embed model error terms as diagnostic instruments in different parts of the model. Identifying the structure of the discrepancy between the two models by model error embedding, can highlight the modeling assumptions that are likely the dominant source of predictive discrepancy. Similarly, the specific form of embedding can be employed as a diagnostic instrument to identify, *e.g.* the quality of one submodel correction versus another.

2.3 Likelihood Construction

In order to find the posterior distribution of the model parameters, we use Bayesian inference to estimate $\tilde{\alpha}$. Given data $\mathcal{D}$, we write Bayes' rule, expressing $p(\tilde{\alpha}|\mathcal{D})$, the posterior density of $\tilde{\alpha}$ conditioned on $\mathcal{D}$, as

$$p(\tilde{\alpha}|\mathcal{D}) = \frac{p(\mathcal{D}|\tilde{\alpha})p(\tilde{\alpha})}{p(\mathcal{D})}, \tag{15}$$

where $p(\mathcal{D}|\tilde{\alpha})$ is the likelihood, $p(\tilde{\alpha})$ is the prior, and $p(\mathcal{D})$ is the evidence which can be treated as a constant in the parameter estimation context. Implicit in the above is also the conditioning on the NL model being fitted, which we leave out for convenience of notation. A key step in obtaining the posterior distribution is the construction of a justifiable likelihood. In general, this is a significant challenge [62,63] where alternate approximations are possible. One convenient approximation, which we choose here, is to use Approximation Bayesian Computation (ABC), a likelihood-free method [76–78] that is often necessary to deal with the challenge of computing expensive/intractable likelihoods. Rather than relying on a likelihood to provide a measure of agreement between model predictions and data, ABC methods rely on a measure of distance between summary statistics evaluated from the two data sources. With the summary statistics on the model output S_h, and those estimated from the data S_q, ABC relies on a kernel density $g(z)$ (a Gaussian), a distance metric $d(S_h, S_q)$, and a tolerance parameter ϵ to provide a pseudo-likelihood:

$$\mathcal{L}(\tilde{\alpha}) = \epsilon^{-1} g(\epsilon^{-1} d(S_h, S_q)) = \frac{1}{\epsilon\sqrt{2\pi}} \exp\left(-\frac{d(S_h, S_q)^2}{2\epsilon^2}\right).$$

Here, for the definition of the distance, we follow [62]. We consider the mean $\mu_j = \mathbb{E}_\xi[h(o_j, \tilde{\alpha}, \xi)]$ and standard deviation $\sigma_j = \sigma_\xi[h(o_j, \tilde{\alpha}, \xi)]$ statistics from the computational model predictions and subtract them to the data q_j and a scaled absolute difference between the mean prediction and the data $\gamma|\mu_j - q_j|$ respectively, where γ is a user-defined parameter. With this, the ABC likelihood reads

$$\mathcal{L}(\tilde{\alpha}) = \frac{1}{\epsilon\sqrt{2\pi}} \prod_{j=1}^{N} \exp(-\frac{(\mu_j - q_j)^2 + (\sigma_j - \gamma|\mu_j - q_j|)^2}{2\epsilon^2}). \tag{16}$$

The motivation for this construction is the desire to require the Bayesian-calibrated model to achieve two goals: (1) fit the data in the mean, and (2) provide a degree of predictive uncertainty that is consistent with the spread of the data around the mean prediction. In particular, this second requirement provides protection against overconfidence in predictions, ensuring that predictive uncertainty is representative of the discrepancy from the data resulting from

model error, irrespective of data size.

3. NONLOCAL OPERATOR REGRESSION WITH EMBEDDED MODEL ERROR

In this section we introduce the embedded treatment of nonlocal operator regression, along with implementation details.

3.1 Mathematical Formulation

We propose the embedded nonlocal operator regression (ENOR) construction, which aims to quantify the model error in nonlocal operator learning using Bayesian inference. In particular, we incorporate a location-dependent GP, $K_\xi(x,\tilde{\omega})$, to represent embedded model error. Here, this Gaussian random field is defined on $\bar{\Omega}\times\Omega_p$, where $\bar{\Omega}$ is the spatial domain together with the nonlocal boundary region, and Ω_p is the sample space of a probability space. The dependence on the GP parameters is implicit, and is suppressed here for convenience of notation. To account for the structural error of learning a homogenized kernel $K_{\mathbf{C}}$, we modify (3) as:

$$\mathcal{L}_{ENL}[u](x,t) := \int_{B_\delta(x)} \left(K_{\mathbf{C}}(|y-x|) \left(1 + K_\xi \left(\frac{x+y}{2}, \tilde{\omega} \right) \right) \right) (u(y,t)-u(x,t))dy. \tag{17}$$

Here the GP $K_\xi(x,\tilde{\omega})$ is defined by a zero mean and the following covariance function:

$$\mathrm{Cov}(K_\xi(x), K_\xi(y)) = \sigma_{gp}^2 \exp\left(-\frac{|x-y|}{l_{gp}} \right),$$

with σ_{gp} and l_{gp} being learnable parameters. Then, the nonlocal model (2) is modified as

$$\frac{\partial^2 u_{ENL}}{\partial t^2}(x,t) - \mathcal{L}_{ENL}[u_{ENL}](x,t) = f(x,t). \tag{18}$$

Note that the corrected kernel preserves the symmetry property in (4)

$$K_{\mathbf{C}}(|y-x|) \left(1 + K_\xi \left(\frac{x+y}{2}, \tilde{\omega} \right) \right) = K_{\mathbf{C}}(|x-y|) \left(1 + K_\xi \left(\frac{y+x}{2}, \tilde{\omega} \right) \right),$$

and correspondingly the fundamental momentum preserving and invariance properties. Here, we allow the embedded model kernel to be location-dependent, to capture the modeling error introduced by the homogenized surrogate with a radial kernel $K_{\mathbf{C}}$. As a result, the corrected kernel is no longer radial, but it satisfies the physical requirement of being

symmetric with respect to x and y .

To represent the GP, we use the Karhunen–Loève expansion [79,80] (KLE), *i.e.*,

$$K_\xi(x,\tilde{\omega}) = \sigma_{gp}\sum_{n=1}^{\infty}\sqrt{\lambda_i}\phi_i(x)\xi_i(\tilde{\omega}), \tag{19}$$

where (λ_i, ϕ_i) are eigenpairs of the kernel function $\exp\left(-\frac{|x-y|}{l_{gp}}\right)$, and ξ_i are independent standard normal random variables. The analytical expression of the eigen-pairs can be found in [81], written for $i = 1, \ldots$ as

$$\begin{aligned}
&\lambda_i = \frac{2/l_{gp}}{(1/l_{gp})^2 + w_i^2},\\
&\phi_i(x) = \tau\left(\cos(w_i x) + \frac{1}{l_{gp}w_i}\sin(w_i x)\right),\\
&\tau = \left\{\frac{1}{2}\left(L(1+(\frac{1}{l_{gp}w_i})^2) + \frac{\sin(2w_iL)}{2w_i}(1-(\frac{1}{l_{gp}w_i})^2) - \frac{1}{l_{gp}w_i^2}(\cos(2w_iL)-1)\right)\right\},
\end{aligned} \tag{20}$$

where L is the length of the domain, τ is a normalizer, and the w_i are obtained by solving the following equation

$$\left(w_i^2 - (1/l_{gp})^2\right)\tan(w_iL) - 2\frac{w_i}{l_{gp}} = 0, \quad i = 1, 2, \ldots . \tag{21}$$

We truncate the summation up to R terms for computational purpose, where R is chosen such that

$$\sum_{i=1}^{R}\lambda_i \geq 0.9\sum_{i=1}^{\infty}\lambda_i.$$

Substituting (19) into (17), we obtain:

$$\frac{\partial^2 u_{ENL}}{\partial t^2}(x,t) - \int_{B_\delta(x)} K_{\mathbf{C}}(|y-x|)(1+\sigma_{gp}\boldsymbol{\Psi}^T\boldsymbol{\xi})(u_{ENL}(y,t)-u_{ENL}(x,t))dy = f(x,t), \tag{22}$$

where

$$\boldsymbol{\Psi}\left(\frac{x+y}{2}\right) := \left[\sqrt{\lambda_1}\phi_1\left(\frac{x+y}{2}\right), ..., \sqrt{\lambda_R}\phi_M\left(\frac{x+y}{2}\right)\right]^T, \quad \boldsymbol{\xi} := [\xi_1, ..., \xi_R]^T.$$

With the KLE we get a realization of the GP by generating a sample of ξ. Then, the numerical scheme in (6) can be employed to evaluate the solution. Denoting $(u^s_{ENL,\mathbf{C}})^{n+1}_{i,k} := u^s_{ENL,\mathbf{C}}(x_i, t^{n+1}, \boldsymbol{\xi}_k)$ as the numerical solution at

(x_i, t^{n+1}) for the s-th sample and k-th GP realization, we have

$$(u^s_{ENL,\mathbf{C}})^{n+1}_{i,k} = 2(u^s_{ENL,\mathbf{C}})^n_{i,k} - (u^s_{ENL,\mathbf{C}})^{n-1}_{i,k} + \Delta t^2 f^s(x_i, t^n) + \Delta t^2 \Delta x \sum_{x_j \in B_\delta(x_i)\cap\chi} K_{\mathbf{C}}(|x_j - x_i|)\left(1 + \sigma_{gp}\mathbf{\Psi}\left(\frac{x_j + x_i}{2}\right)^T \boldsymbol{\xi}_k\right)((u^s_{ENL,\mathbf{C}})^n_{j,k} - (u^s_{ENL,\mathbf{C}})^n_{i,k}), \tag{23}$$

where $\boldsymbol{\xi}_k$ is the k-th realization of the random vector generating the corresponding GP. Similarly, the step-wise version of the embedded nonlocal model can be written as

$$(\tilde{u}^s_{ENL,\mathbf{C}})^{n+1}_{i,k} = 2(u^s_{DNS})^n_i - (u^s_{DNS})^{n-1}_i + \Delta t^2 f^s(x_i, t^n) + \Delta t^2 \Delta x \sum_{x_j \in B_\delta(x_i)\cap\chi} K_{\mathbf{C}}(|x_j - x_i|)\left(1 + \sigma_{gp}\mathbf{\Psi}\left(\frac{x_j + x_i}{2}\right)^T \boldsymbol{\xi}_k\right)((u^s_{DNS})^n_j - (u^s_{DNS})^n_i). \tag{24}$$

To effectively evaluate the pseudo-likelihood, we employ the expression (16). Note that the generated eigenpairs (λ_i, ϕ_i) are dependent on l_{gp}. Since optimizing with respect to l_{gp} together with the other parameters is computationally infeasible, we treat l_{gp} as a tunable hyperparameter, and perform our ENOR algorithm for a fixed l_{gp} at a time to avoid the repeated cost of (20) and (21). Therefore the enhanced parameter set $\tilde{\alpha} = (\mathbf{C}, \sigma_{gp}, l_{gp})$ will be reduced to $\tilde{\alpha} = (\mathbf{C}, \sigma_{gp})$. For each observation, we have the following ABC likelihood

$$\begin{aligned}\mathcal{L}(\tilde{\alpha}) &:= p(\mathcal{D}|(\mathbf{C}, \sigma_{gp})) \\ &= \prod_{s,i,n=1}^{S,L,T/\Delta t} \exp\left(-\frac{1}{2\epsilon^2}\left((\mu^s_{\mathbf{C}}(x_i, t^n) - u^s_{DNS}(x_i, t^n))^2 + (\sigma^s_{\mathbf{C}}(x_i, t^n) - \gamma|\mu^s_{\mathbf{C}}(x_i, t^n) - u^s_{DNS}(x_i, t^n)|)^2\right)\right),\end{aligned} \tag{25}$$

where

$$\mu^s_{\mathbf{C}}(x_i, t^n) = \frac{1}{K}\sum_{k=1}^K (u^s_{ENL,\mathbf{C}})^n_{i,k}, \tag{26}$$

$$\sigma^s_{\mathbf{C}}(x_i, t^n) = \sqrt{\frac{1}{K-1}\sum_{k=1}^K ((u^s_{ENL,\mathbf{C}})^n_{i,k} - \mu^s_{\mathbf{C}}(x_i, t^n))^2} \tag{27}$$

are the sample mean and standard deviation for $(u^s_{ENL,\mathbf{C}})^n_i$ and $(u^s_{ENL,\mathbf{C}})^n_i$, for K samples of ξ, and with the current value of σ_{gp}.

In [47], we found that a good prior distribution plays a critical role in achieving a fast convergence of the MCMC

algorithm and we used the learnt kernel parameter $\mathbf{C}_0$ from a deterministic nonlocal operator regression (DNOR) to construct such a prior. To provide a prior on $\mathbf{C}$, we use independent standard normal priors on the kernel parameters $C_m \sim \mathcal{N}(C_{0,m}, \hat{\sigma}^2)$, $m = 1, ...M-2$, with $\mathbf{C}_0$ being the learnt kernel parameter from DNOR, and the standard deviation $\hat{\sigma}$ as a tunable hyperparameter. In practice, since $\sigma_{gp} > 0$, we infer $\ln(\sigma_{gp})$ instead of σ_{gp}. To get an initial state for $\ln(\sigma_{gp})$, we optimize the following

$$\ln(\sigma_{gp,0}) := \underset{\ln(\sigma_{gp})}{\operatorname{argmin}} \frac{1}{2\epsilon^2} \sum_{s,i,n=1}^{S,L,T/\Delta t} (\mu^s_{\mathbf{C}}(x_i,t^n) - u^s_{DNS}(x_i,t^n))^2 + (\sigma^s_{\mathbf{C}}(x_i,t^n) - \gamma|\mu^s_{\mathbf{C}}(x_i,t^n) - u^s_{DNS}(x_i,t^n))|^2) \tag{28}$$

and assign a uniform prior on $\ln(\sigma_{gp})$, specifically $\ln(\sigma_{gp}) \sim \mathcal{U}[\ln(\sigma^{lo}_{gp}), \ln(\sigma^{hi}_{gp})]$. We treat σ^{lo}_{gp} and σ^{hi}_{gp}, the lower and upper bounds on σ_{gp}, as tunable hyperparameters. Once the proposal for $\ln(\sigma_{gp})$ exceeds the bounds, the log-posterior will be set to $-\infty$ automatically. Combining the likelihood in (25) and the prior (with $\ln(\sigma_{gp}) \in [\ln(\sigma^{lo}_{gp}), \ln(\sigma^{hi}_{gp})]$), we can finally define the unnormalized posterior $p(\mathbf{C}, \sigma_{gp}|\mathcal{D}) \propto p(\mathcal{D}|\mathbf{C}, \sigma_{gp})p(\mathbf{C}, \sigma_{gp}))$ and obtain the negative log-posterior after eliminating the constant terms:

$$\frac{1}{2\epsilon^2} \sum_{s,i,n=1}^{S,L,T/\Delta t} (\mu^s_{\mathbf{C}}(x_i,t^n) - u^s_{DNS}(x_i,t^n))^2 + (\sigma^s_{\mathbf{C}}(x_i,t^n) - \gamma|\mu^s_{\mathbf{C}}(x_i,t^n) - u^s_{DNS}(x_i,t^n))|^2) + \frac{||\mathbf{C}-\mathbf{C}_0||^2_{l_2}}{2\hat{\sigma}^2}. \tag{29}$$

3.2 Implementation Details

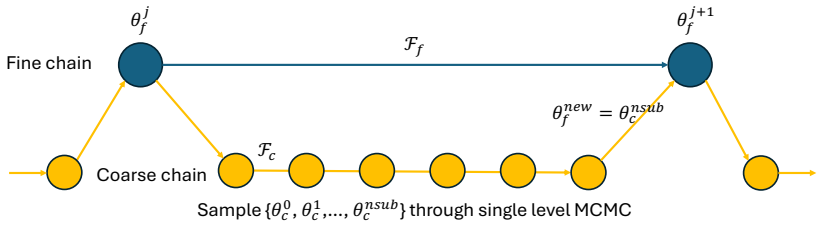


Fig. 1. Schematic of generating a proposal θ' for a two-level MLDA algorithm.

As can be seen from (29), the evaluation of the posterior requires the estimation of the first and second order moments of the output using K GP samples. For each sample, the nonlocal meshfree method (23) is applied for each spatial point, time step, and observation, making this numerical evaluation expensive in MCMC. In order to improve the efficiency of the MCMC procedure, we employ the Multilevel Delayed Acceptance (MLDA) MCMC technique

[64,65], which exploits a hierarchy of models of increasing complexity to efficiently generate samples from an unnormalized target distribution. For the purpose of illustration, we summarize the key factors of two-level Delayed Acceptance (TLDA) MCMC here, while the method could be extended to a model hierarchy with arbitrarily many levels by recursion. For the vanilla Metropolis-Hastings (MH) algorithm which is a typical single level MCMC, consider sampling a trace $\{\theta^1, ..., \theta^{n_{single}}\}$ from a target distribution $\pi_t(\cdot)$, using a proposal distribution $\pi_p(\cdot|\cdot)$, and an initial state θ^0. MH accepts a new proposal θ^{new} given θ^j for $j = 0, 1, ...n_{single} - 1$ with probability

$$\min\left\{1, \frac{\pi_t(\theta^{new})\pi_p(\theta^j|\theta^{new})}{\pi_t(\theta^j)\pi_p(\theta^{new}|\theta^j)}\right\},$$

otherwise it rejects θ^{new} and sets $\theta^{j+1} = \theta^j$. In the context of Bayesian inference, the MH target distribution is the posterior distribution, i.e. $\pi_t(\theta|\mathcal{D}) = p(\theta|\mathcal{D})$. Unlike the vanilla MCMC which has only one model, in TLDA, a cheaper model is employed to reduce the computational cost. Fig.1 illustrates the work flow for TLDA, where the coarse subchains are sampled in order to provide proposals for the fine model. Denote by $\mathcal{F}_f$ the fine forward model and by $\mathcal{F}_c$ the coarse forward model, with π_f and π_c being their target distributions respectively. Starting from θ_f^j, in the coarse level, one can generate the subchain $\{\theta_c^1, \theta_c^2, ..., \theta_c^{n_{sub}}\}$ of length n_{sub} using MH or any single-level MCMC method. After the subchain is finished, we take $\theta_c^{n_{sub}}$ as the proposal for the fine chain (i.e. $\theta_f^{new} = \theta_c^{n_{sub}}$) and accept this proposal with probability

$$\min\left\{1, \frac{\pi_f(\theta_f^{new})\pi_c(\theta_f^j)}{\pi_f(\theta_f^j)\pi_c(\theta_f^{new})}\right\}, \tag{30}$$

otherwise reject θ_f^{new} and set $\theta_f^{j+1} = \theta_f^j$ [65].

When the approximation provided by the coarse model is poor, many samples will be rejected by the fine model resulting in a very low acceptance rate. As outlined in [65], an enhanced Adaptive Error Model (AEM) based on [82] is useful to account for and correct the discrepancy between the fine and coarse models. We use the two-level AEM [65,82] in the present use of TLDA. For parameters $\tilde{\alpha}$ and operating conditions o, the bias $\mathcal{B}(o)$ between the two models can be written as

$$\mathcal{B}(o) = \mathcal{F}_f(o, \tilde{\alpha}, \xi) - \mathcal{F}_c(o, \tilde{\alpha}, \xi) \tag{31}$$

When the parameter set $\tilde{\alpha}$ is sampled from the prior distribution, then

$$\mathcal{B}(o) \sim \mathcal{F}_f(o, \tilde{\alpha}, \xi) - \mathcal{F}_c(o, \tilde{\alpha}, \xi).$$

Denoting the fine model solution as $u^s_{ENL,\mathbf{C}}(x_i,t^n,\boldsymbol{\xi}) := u^s_{ENL}(\mathbf{C},\ln(\sigma_{gp}),\boldsymbol{\xi})^n_i$, the coarse model (the stepwise nonlocal model) solution as $\tilde{u}^s_{ENL,\mathbf{C}}(x_i,t^n,\boldsymbol{\xi}) := \tilde{u}^s_{ENL}(\mathbf{C},\ln(\sigma_{gp}),\boldsymbol{\xi})^n_i$, and the trainable parameter set as $\tilde{\beta} := (\mathbf{C},\ln(\sigma_{gp}),\boldsymbol{\xi})$, we have the following sample mean and standard deviation for the correction term

$$\hat{\mathbb{E}}[B^{s,n}_i] = \mu^s(x_i,t^n) - \tilde{\mu}^s(x_i,t^n), \quad (\hat{\sigma}[B^{s,n}_i])^2 = (\sigma^s(x_i,t^n))^2 - (\tilde{\sigma}^s(x_i,t^n))^2, \tag{32}$$

where

$$\begin{aligned}
\mu^s(x_i,t^n) &= \hat{\mathbb{E}}_{\tilde{\beta}}[u^s_{ENL}(\tilde{\beta})^n_i] = \frac{1}{N_0}\sum_{n_0=1}^{N_0} u^s_{ENL}(\tilde{\beta}_{n_0})^n_i,\\
\sigma^s(x_i,t^n) &= \hat{\mathbb{V}}^{1/2}_{\tilde{\beta}}[u^s_{ENL}(\tilde{\beta})^n_i] = \sqrt{\frac{1}{N_0-1}\sum_{n_0=1}^{N_0}(u^s_{ENL}(\tilde{\beta}_{n_0})^n_i - \mu^s(x_i,t^n))^2},\\
\tilde{\mu}^s(x_i,t^n) &= \hat{\mathbb{E}}_{\tilde{\beta}}[\tilde{u}^s_{ENL}(\tilde{\beta})^n_i] = \frac{1}{N_0}\sum_{n_0=1}^{N_0} \tilde{u}^s_{ENL}(\tilde{\beta}_{n_0})^n_i,\\
\tilde{\sigma}^s(x_i,t^n) &= \hat{\mathbb{V}}^{1/2}_{\tilde{\beta}}[\tilde{u}^s_{ENL}(\tilde{\beta})^n_i] = \sqrt{\frac{1}{N_0-1}\sum_{n_0=1}^{N_0}(\tilde{u}^s_{ENL}(\tilde{\beta}_{n_0})^n_i - \tilde{\mu}^s(x_i,t^n))^2}.
\end{aligned} \tag{33}$$

Here, N_0 is the number of samples of $\tilde{\beta}$ which will be used for computing the sample mean and standard deviation. We highlight that the enhanced model parameter set $\tilde{\alpha} = (\mathbf{C},\ln(\sigma_{gp}))$ should be sampled simultaneously with $\boldsymbol{\xi}$. In contrast, μ and σ are calculated by averaging over multiple samples of $\boldsymbol{\xi}$ only, for each fixed enhanced parameter set $\tilde{\alpha} = (\mathbf{C},\ln(\sigma_{gp}))$ in (26). In other words, the sample moments computed in (33) are to be used for correction for any parameter $\tilde{\alpha}$ inside the prior distribution instead of a fixed parameter.

By using the AEM technique, (29) can be well approximated by the coarse model following

$$\begin{aligned}
&\frac{1}{2\epsilon^2}\sum_{s,i,n=1}^{S,L,T/\Delta t}(\tilde{\mu}^s_{\mathbf{C}}(x_i,t^n) + \hat{\mathbb{E}}(\mathcal{B}^{s,n}_i) - u^s_{DNS}(x_i,t^n))^2\\
&+ (\sqrt{(\tilde{\sigma}^s_{\mathbf{C}}(x_i,t^n))^2 + \hat{\mathbb{V}}(B^{s,n}_i)} - \gamma|\tilde{\mu}^s_{\mathbf{C}}(x_i,t^n) + \hat{\mathbb{E}}(\mathcal{B}^{s,n}_i) - u^s_{DNS}(x_i,t^n))^2|) + \frac{||\mathbf{C}-\mathbf{C}_0||^2_{l_2}}{2\hat{\sigma}^2},
\end{aligned} \tag{34}$$

where $\hat{\mathbb{E}}(\mathcal{B}^{s,n}_i)$ and $\hat{\mathbb{V}}(\mathcal{B}^{s,n}_i)$ are the sample mean and variance for the correction term $\mathcal{B}$ computed using (33). Technically, this step is also a reference for tuning the upper and lower bounds for the prior distribution of $\ln(\sigma_{gp})$. Such an interval should be selected in such a way that at least at $\mathbf{C} = \mathbf{C}_0$ the log-negative posterior in (34) which is approximated by the coarse model and the AEM could roughly reproduce the log-negative posterior in (29) which is

generated using the fine model. We summarize our method in Algorithm 1.

Algorithm 1: A Two-Phase Learning Algorithm

1: **Find a good initial state**

1a) Learn the estimated kernel parameter, $\mathbf{C}_0 = C_{0,1}, ..., C_{0,M-2}$, by minimizing the time-accumulated error via

$$\mathbf{C}_0 := \underset{\mathbf{C}}{\text{argmin}} \sum_{s=1}^{S} \frac{||u^s_{NL,\mathbf{C}} - u^s_{DNS}||^2_{l^2(\Omega\times[0,T])}}{||u^s_{DNS}||^2_{l^2(\Omega\times[0,T])}} + \lambda||\mathbf{C}||^2_{l^2}. \tag{35}$$

1b) With fixed $\mathbf{C} = \mathbf{C}_0$, find $\ln(\sigma_{gp,0})$ by minimizing

$$\ln(\sigma_{gp,0}) := \underset{\ln(\sigma_{gp})}{\text{argmin}} \sum_{s,i,n=1}^{S,L,T/\Delta t} ((\mu^s_{\mathbf{C}}(x_i,t^n) - u^s_{DNS}(x_i,t^n))^2 + (\sigma^s_{\mathbf{C}}(x_i,t^n) - \gamma|\mu^s_{\mathbf{C}}(x_i,t^n) - u^s_{DNS}(x_i,t^n)|)^2). \tag{36}$$

2: **Perform MLDA**

2a) Using prior $C_m \sim N(C_{0,m}, \hat{\sigma}^2)$, $m = 0, ..., M-2$ and $\ln(\sigma_{gp}) \sim \mathcal{U}[\ln(\sigma^{lo}_{gp}), \ln(\sigma^{hi}_{gp})]$, compute the sample mean and standard deviation of the correction term $\mathcal{B}$ following

$$\hat{\mathbb{E}}[\mathcal{B}^{s,n}_i] = \mu^s(x_i,t^n) - \tilde{\mu}^s(x_i,t^n), \quad \hat{\mathbb{V}}[\mathcal{B}^{s,n}_i] = (\sigma^s(x_i,t_n))^2 - (\tilde{\sigma}^s(x_i,t^n))^2$$

and tune $[\ln(\sigma^{lo}_{gp}), \ln(\sigma^{hi}_{gp})]$ by recursively doing this step until the loss given by (34) could roughly match the loss given by (29) at $\mathbf{C} = \mathbf{C}_0$ inside the interval.

2b) Perform MLDA by evaluating the following negative log-posterior in the coarse level

$$\begin{aligned}\frac{1}{2\epsilon^2} \sum_{s,i,n=1}^{S,L,T/\Delta t} &(\tilde{\mu}^s_{\mathbf{C}}(x_i,t^n) + \hat{\mathbb{E}}(\mathcal{B}^{s,n}_i) - u^s_{DNS}(x_i,t^n))^2 + (\sqrt{(\sigma^s_{\mathbf{C}}(x_i,t^n))^2 + \hat{\mathbb{V}}[\mathcal{B}^{s,n}_i]} \\ &- \gamma|\mu^s_{\mathbf{C}}(x_i,t^n) + \hat{\mathbb{E}}(\mathcal{B}^{s,n}_i) - u^s_{DNS}(x_i,t^n))^2|) + \frac{||\mathbf{C}-\mathbf{C}_0||^2_{l_2}}{2\hat{\sigma}^2},\end{aligned} \tag{37}$$

and the following negative log-posterior in the fine level

$$\frac{1}{2\epsilon^2} \sum_{s,i,n=1}^{S,L,T/\Delta t} (\mu^s_{\mathbf{C}}(x_i,t^n) - u^s_{DNS}(x_i,t^n))^2 + (\sigma^s_{\mathbf{C}}(x_i,t^n) - \gamma|\mu^s_{\mathbf{C}}(x_i,t^n) - u^s_{DNS}(x_i,t^n))|^2) + \frac{||\mathbf{C}-\mathbf{C}_0||^2_{l_2}}{2\hat{\sigma}^2}. \tag{38}$$

3: **Postprocessing**

Perform convergence check for the parallel chains, extract effective samples from the MCMC chain, analyze the uncertainty of the corresponding solution and other quantities of interests.

4. APPLICATION: HOMOGENIZATION FOR A HETEROGENEOUS ELASTIC BAR

In this section, we examine the efficacy of the proposed ENOR approach on inferring the nonlocal homogenized surrogate for modeling the propagation of stress waves through a one-dimensional bar [22,47]. In particular, we consider

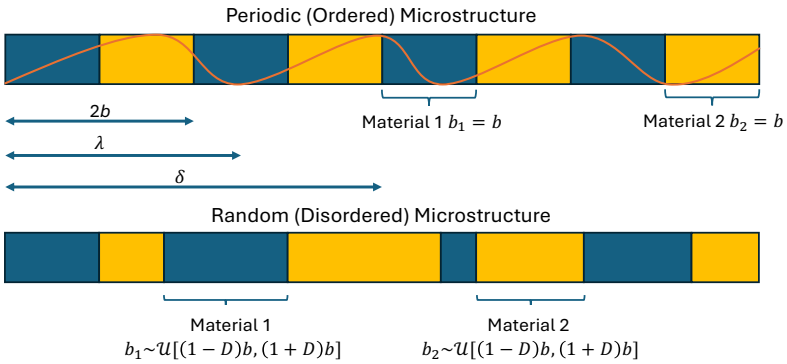


Fig. 2. One-dimensional bar composite of material 1 and material 2. (Top) periodic microstructure with fixed layer size$= b$. (Bottom) Random microstructure with layer size satisfying random distribution$\sim \mathcal{U}[(1-D)b, (1+D)b]$.

a composite bar either made of periodic layers or randomly generated layers (see Fig.2), and we assume that the amplitude of the waves is sufficiently small, so that a linear elasticity model is a valid way to describe the wave motion within the layers and at their interfaces. With this assumption of linear elastodynamics, the propagation of waves through the bar can be described as:

$$\frac{\partial^2 u}{\partial t^2}(x,t) - \frac{1}{\rho}\frac{\partial}{\partial x}\left(E(x)\frac{\partial u}{\partial x}(x,t)\right) = \frac{1}{\rho}f(x,t), \tag{39}$$

where ρ is the mass density which is assumed to be constant throughout the body, u is the displacement, and f is the external load density. $E(x)$ is the elastic modulus which varies spatially according to the microstructure, *i.e.*, we have $E = E_1$ in the blue regions of Fig.2, and $E = E_2$ in the yellow regions. $s(x,t) := E(x)\frac{\partial u}{\partial x}(x,t)$ is the stress. On the interfaces of two materials, the following jump conditions hold: $[u(x,t)] = 0$, $\left[E(x)\dfrac{\partial u}{\partial x}(x,t)\right] = 0$.

Ideally, one can solve (39) using numerical solvers on fine discretizations of the computational domain to explicitly represent all interfaces. However, in real-world applications such as projectile impact modeling [21,83], one is interested in modeling the decay of the wave over distances that are several thousand times larger than the layer size. In these circumstances, a fine numerical solver is prohibitively expensive and a homogenized surrogate model is desired to provide scalable predictions.

In this example, we first generate short-term high-fidelity simulation data by solving (39) using characteristic line method. This method, which we denote as the direct numerical simulation (DNS) technique, assumes that the waves running in the opposite direction converge on the node, and update the material velocity explicitly from the jumping condition which is a consequence of the momentum conservation. Due to this property, this DNS solver is free of truncation error and approximation error as in the classical PDE solver, which allow us to simulate the exact velocity of wave propagation through arbitrarily many microstructural interfaces. We refer the reader to [21,47] for more details

of this method for further details. Then, our goal is to construct a nonlocal homogenized surrogate model from this high-fidelity simulation dataset $\mathcal{D}$.

4.1 Example 1: Periodic Microstructure

Throughout the section, a non-dimensionalized setting is employed for the physical quantities for the simplicity of numerical experiments, following the setup in [21]. First, we consider a periodic heterogeneous bar, where the layer size for the two materials is a constant $b = 0.2$. The bar length is $L = 20$ and the physical domain is set to be $[-L/2, L/2]$. Components 1 and 2 have the same density $\rho = 1$ and Young's moduli are set as $E_1 = 1$ and $E_2 = 0.25$, respectively. For the purpose of training and validation, three types of datasets/settings are considered:

Setting 1: Oscillating source (20 loading instances). We set $L = 20$. The bar starts from rest such that $v(x, 0) = u(x, 0) = 0$, and an oscillating loading is applied with $f(x,t) = e^{-\left(\frac{2x}{5kb}\right)^2} e^{-\left(\frac{t-t_0}{t_p}\right)^2} \cos^2\left(\frac{2\pi x}{kb}\right)$ with $k = 1, 2, \ldots, 20$. Here we take $t_0 = t_p = 0.8$.

Setting 2: Plane wave with ramp (11 loading instances). We also set the domain parameter as $L = 20$. The bar starts from rest ($u(x, 0) = 0$) and is subject to zero loading ($f(x, t) = 0$). For the velocity on the left end of the bar, we prescribe

$$v(-L/2, t) = \begin{cases} \sin(\omega t)\sin^2\left(\dfrac{\pi t}{30}\right), & t \le 15 \\ \sin(\omega t), & t > 15 \end{cases}$$

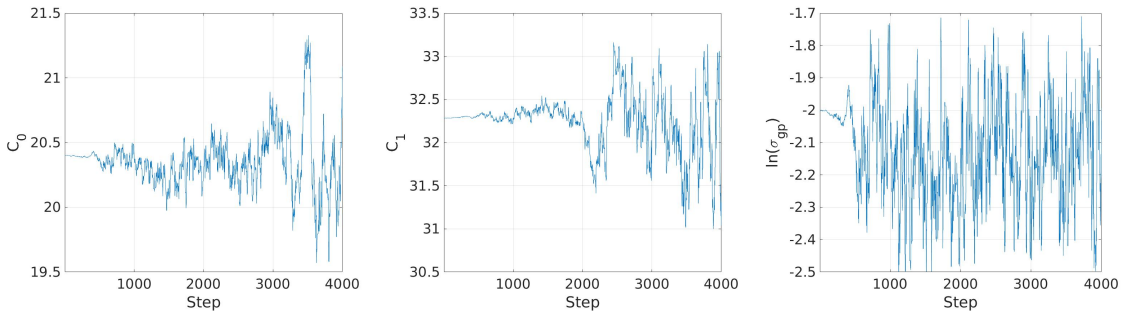
for $\omega = 0.35, 0.7, \cdots, 3.85$.

Setting 3: Wave packet (3 loading instances). We consider a longer bar with $L = 266.6$, with the bar starting from rest ($u(x, 0) = 0$), and is subject to zero loading ($f(x, t) = 0$). The velocity on the left end of the bar is prescribed as $v(-L/2, t) = \sin(\omega t)\exp\left(-(t/5 - 3)^2\right)$ with $\omega = 2$, 3.9, and 5.

For all data types, the parameters for the nonlocal solver and the optimization algorithm are set to $\Delta x = 0.05$, $\Delta t = 0.02$, $\delta = 1.2$, and $M = 24$. For training purposes, we generate data of types 1 and 2 till $T = 2$. Then, to investigate the performance of our surrogate model in long-term prediction tasks we simulate till $T = 100$ for setting 3. We choose $\hat{\sigma}^2 = \frac{0.01^2}{0.35}$ and $[\ln(\sigma_{gp}^{lo}), \ln(\sigma_{gp}^{hi})] = [-2.5, -1.7]$ to achieve optimal performance. A detailed ablation study of these parameters is provided in Appendix A.

4.1.1 Results from MCMC Experiments

We use PyMC [84] for all the MCMC computations below. We begin by verifying the convergence behavior of our algorithm. In particular, we test a single-level MCMC with the differential evolution Metropolis (DEMetropolis) algorithm [85] on the fine level with 4,000 draws and a burn-in stage of 300. DEMetropolis, or DEMetropolis(Z) is a variation of Metropolis-Hastings algorithm that uses randomly selected draws from the past to make more educated jumps. Results in Fig. 3 indicate that the single-level MCMC suffers from an extremely long burn-in stage and poor mixing. We treat these single-level results as a baseline which we use to examine the relative performance of the multilevel method below.

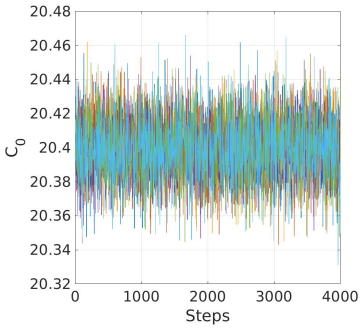


(a) First parameter (b) Second parameter (c) $\ln(\sigma_{gp})$

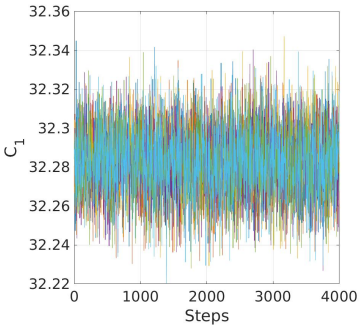
Fig. 3. Trace plot for the MCMC using single-level DEMetropolisZ sampler.

We first illustrate the utility of the multilevel algorithm using a correlation length $l_{gp} = L/2$. We run 6 independent chains, each with 4,000 draws, a burn-in stage of 300 and a subchain length on the coarse level of 100. Each chain is initialized using the scheme proposed in Section 3. We examine the quality of the chains both visually and quantitatively. Fig.4 illustrates that all the six independent chains attain essentially the same stationary state. Further, using the improved $\hat{R}$ statistic [86] relying on ArviZ in Python, we find that the $\hat{R}$ values for all 24 parameters are very close to 1, with the highest being 1.0019, again highlighting the convergence of the chains. Note that, for an ergodic process, this statistic decays to 1 in the limit of infinite chain length [86].

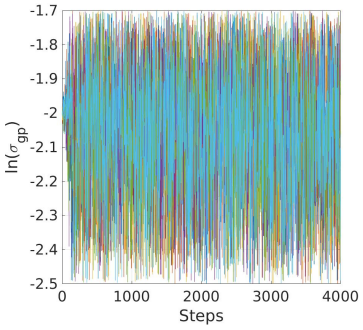
The chains have 24,000 draws in total, with approximately 42% acceptance rate on average. To present the aggregate results, we use roughly 6,000 equally spaced samples out of the total, where this effective sample size (ESS) was calculated using the method in [87]. The trace plots shown in Fig.5 indicate good mixing, and, with the high ESS, we have reliable probability density functions (PDFs) and associated statistics.



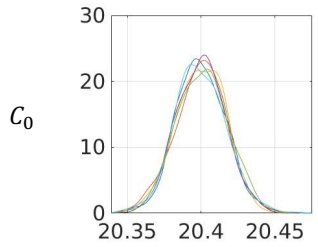
(a) First parameter

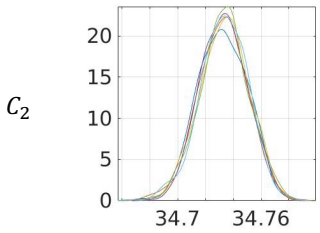


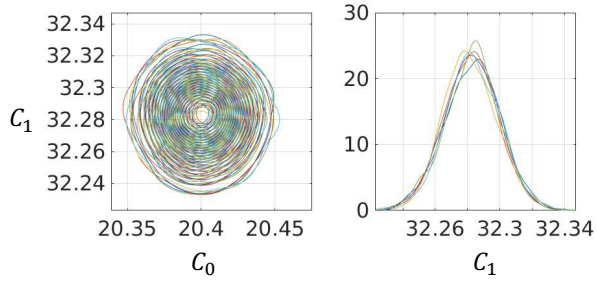
(b) Second parameter



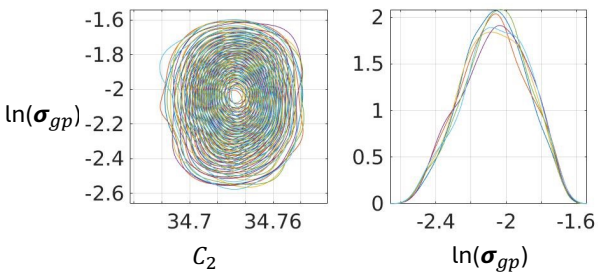
(c) $\ln(\sigma_{gp})$





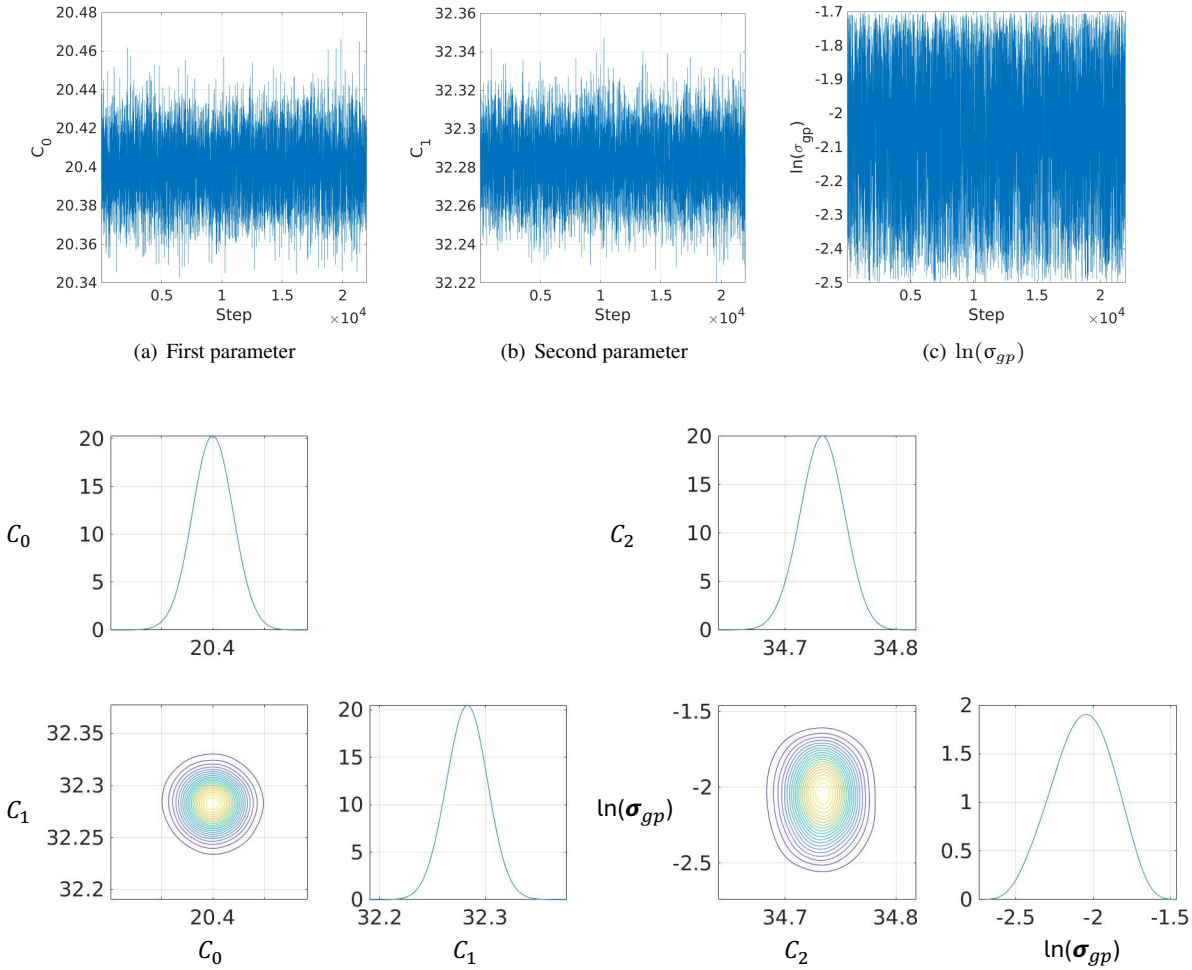


(d) Joint distribution of the first and second parameters in $\mathbf{C}$



(e) Joint distribution of the third parameter in $\mathbf{C}$ and $\ln(\sigma_{gp})$

Fig. 4. Convergence check: Trace plots and PDFs for the traces. For each trace, we have an acceptance rate ≈ 0.42, and an effective sample size $\approx$ 1,000 (out of 4,000 draws).



(a) First parameter (b) Second parameter (c) $\ln(\sigma_{gp})$

(d) Joint distribution of the first and second parameters in $\mathbf{C}$ (e) Joint distribution of of the third parameter in $\mathbf{C}$ and $\ln(\sigma_{gp})$

Fig. 5. Trace plot and PDF for the combined trace. The acceptance rate ≈ 0.42, ESS $\approx$ 6,000 (out of 24,000 draws).

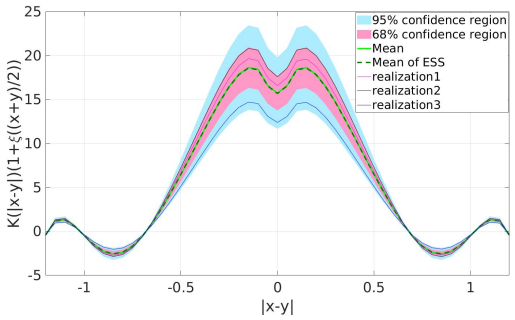
4.1.2 Impact of GP Correlation Lengths

As discussed in Section 3.1, to avoid the repeated cost of (20) and (21) at each MCMC step, we use the GP correlation length, l_{gp}, as a tunable hyperparameter. In this section, we investigate the impact of different values of l_{gp} on the learnt corrected kernels and the corresponding nonlocal solution behaviors.

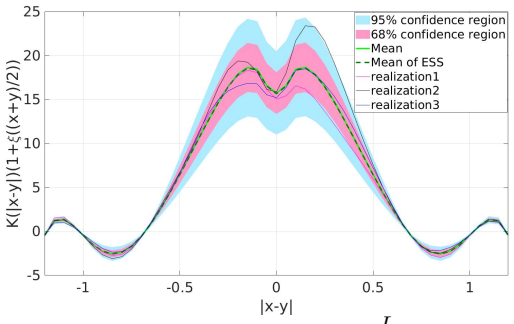
In Fig. 6, the uncertainty with several realizations for the kernel with GP, $K(|x-y|)(1+\xi(\frac{x+y}{2}))$, is plotted for correlation lengths $l_{gp} = 2L = 40$ and $l_{gp} = L/128 = 0.15625$, using 10,000 realizations in total. Since the behavior of the kernel is similar at each point in the one-dimensional bar, we pick the location $x = 0.0$ as an instance for illustration. We observe that the curve labeled 'Mean', which is the mean value of all the realizations, is consistent with the corresponding kernel as the mean of the ESS, which is a single push-forward of the embedded model (18) using the mean value of the effective samples and setting the GP to 0. In fact, without the GP the embedded model degenerates to the original nonlocal model (2). This consistency is because of the linearity of the nonlocal kernel and the independence of the ξ components. Further, for different l_{gp} values, one can barely see changes in the mean and the confidence region. On the other hand, sampled GP realizations exhibit significant structural differences between the two cases, in Figs. 6(a) and 6(b), consistent with the large change in l_{gp} between the two cases.

To examine the dispersion behavior of these kernel realizations, we plot in Figs. 7(a,c) the group velocity of kernels corresponding to a GP with a large ($l_{gp} = 40$) and relatively small ($l_{gp} = 0.625$) correlation length. For each realization of the kernel, the group velocity is computed using a wave packet that travels a long enough distance such that the wave is away from the ends of the bar. In order to do this in numerical simulation, we set the bar length equal to 400 and generate realizations of the kernels with the same correlation length on this longer bar. We note that the reduction in l_{gp} results in a smaller band stop location (the smallest frequency where the group velocity drops to zero). Further, it is evident that the confidence region matches the dispersion behavior better when using a large correlation length. At the same time, oscillation is observed in the low frequency ($\omega < 2.0$) region in the small l_{gp} case. Finally, in Figs. 7(b,d), we plot the dispersion curves of these kernels. Note that all learnt kernels are positive, highlighting the physical stability of the corresponding nonlocal models.

Next, we investigate the posterior uncertainty on the predicted displacements from the learnt models. Two samples from setting 1 and 2 datasets are considered, with forcing at $k = 10$ in setting 1 and $\omega = 1.05$ in setting 2. Let us first recall the definitions of two relevant push-forward posteriors for a typical additive data noise problem setup $y = f_d(x, \lambda, \sigma) = f(x, \lambda) + \epsilon(\sigma)$ given the parameter posterior $p(\lambda, \sigma|D)$. We have the push-forward posterior (PFP) as the push-forward of the parameter posterior through the predictive model $f(x, \lambda)$, while the posterior predictive (PP) is defined as the push-forward of the parameter posterior through the full data model $f_d(x, \lambda, \sigma)$ [88,89]. Note

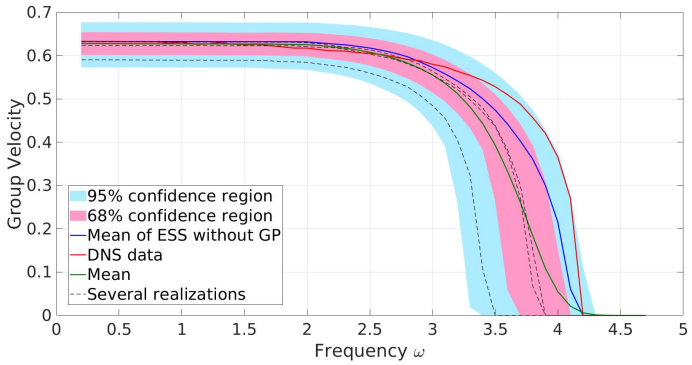


(a) Kernel at point $x = 0.0$ for $l_{gp} = 2L$=40.

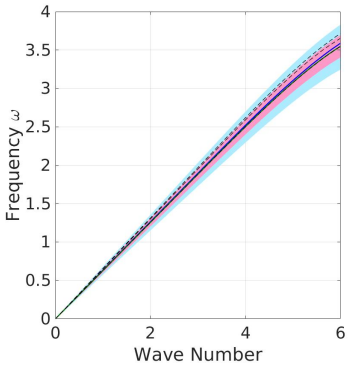


(b) Kernel at point $x = 0.0$ for $l_{gp} = \frac{L}{128}$=0.15625.

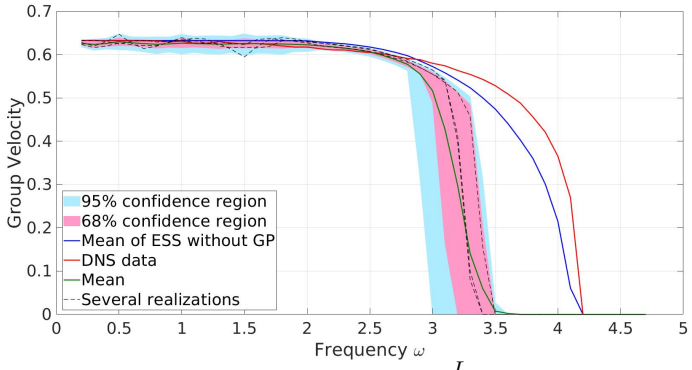
Fig. 6. Kernel with uncertainty for a bar with periodic microstructure using two different correlation length.



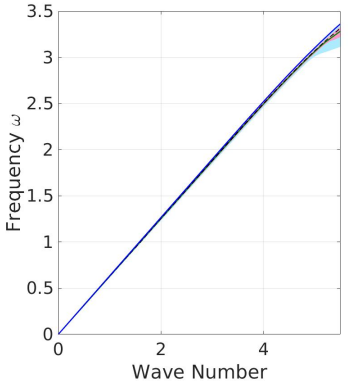
(a) Group velocity using $l_{gp} = 2L$=40.



(b) Dispersion curve using $l_{gp} = 2L$=40.



(c) Group velocity using $l_{gp} = \frac{L}{32}$=0.625.



(d) Dispersion curve using $l_{gp} = \frac{L}{32}$=0.625.

Fig. 7. Group velocity and dispersion curve with uncertainty for a bar with periodic microstructure using two different correlation lengths. The curve labeled 'Mean of ESS without GP' denotes the single push-forward with the GP set to 0 and using the mean value of the effective samples. The curve labeled 'Mean' denotes the mean value of all the realizations.

that, in an embedded model-error construction with no additive data noise, the PFP and PP are equivalent. In Fig.8, the PFP, which is the push-forward of the posterior $p(\mathbf{C}, \sigma_{gp}|\mathcal{D})$ through the nonlocal model with embedded model error, is plotted for two different l_{gp} values. For each case, we see that both the 95% and 68% PFP confidence regions generally cover the majority of the ground truth data over the spatial domain. However, we do observe that, visually, and for these two samples, the kernel with a smaller l_{gp} does a better job of spanning the ground truth data discrepancy from the mean model. This is also illustrated in Fig.10, where the best choices of l_{gp} for different training samples are plotted according to the Continuous Ranked Probability Score (CRPS) [90]. The CRPS compares a single ground truth value to a distribution. Assume that we have a ground truth $\bar{y}$ and a cumulative distribution function (CDF) H for a variable $\bar{x}$, then the CRPS can be analytically written as

$$\text{CRPS}(H, \bar{y}) = \int (H(\bar{x}) - 1_{\{\bar{x} \geq \bar{y}\}})^2 \mathrm{d}\bar{x}.$$

In a numerical setting where only a sampling-based empirical CDF is available, [90] provides alternative forms of the CRPS which are feasible to estimate

$$\text{CRPS}(H, \bar{y}) = \mathbb{E}[|\bar{X} - \bar{y}|] - \frac{1}{2}\mathbb{E}[|\bar{X} - \bar{X}'|] \tag{40}$$

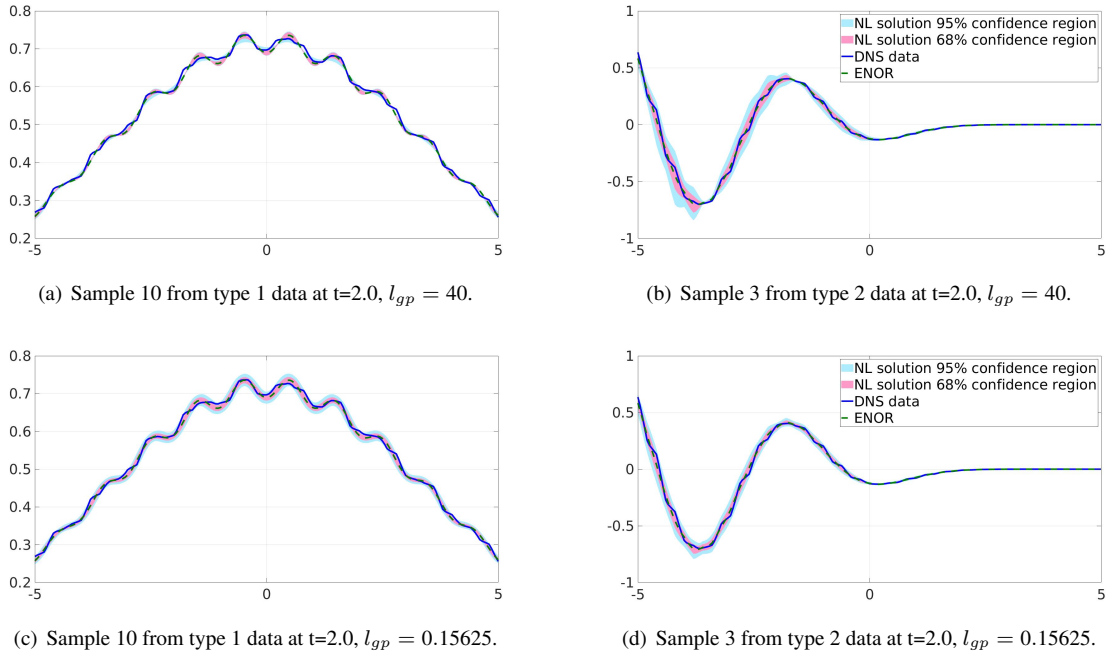
$$= \mathbb{E}[|\bar{X} - \bar{y}|] + \mathbb{E}[\bar{X}] - 2\mathbb{E}[\bar{X} \cdot H(\bar{X})], \tag{41}$$

where $\bar{X}, \bar{X}'$ are independently and identically distributed according to H. Per the definition of the CRPS, the lower the score is, the better does our predicted displacement match the DNS data in distribution. Specifically, we use equation (40) here, but in principle the two expressions are equivalent. The average CRPS values across different training samples, at $t = 2.0$, are summarized in Table.1. As the correlation length l_{gp} decreases, only a slight reduction in the CRPS is obtained, suggesting that all studied kernels have comparable performance in obtaining the correct distribution of the displacement.

Material	$2L$	L	$L/2$	$L/4$	$L/8$	$L/16$	$L/32$	$L/64$	$L/128$
Periodic	0.0048	0.0048	0.0048	0.0048	0.0047	0.0046	0.0045	0.0044	0.0043
Disorder	0.0080	0.0080	0.0080	0.0079	0.0079	0.0077	0.0076	0.0076	0.0076

TABLE 1. Average CRPS of training samples for different l_{gp}. The value of the CRPS is evaluated for all the training samples on each grid in the physical domain at the last time step $t = 2.0$.

Finally, we provide the prediction of a wave packet, a waveform that is substantially different from the training data. With this setting, we consider an extrapolation scenario: the learnt model is employed to generate a long-term



(a) Sample 10 from type 1 data at t=2.0, $l_{gp} = 40$.

(b) Sample 3 from type 2 data at t=2.0, $l_{gp} = 40$.

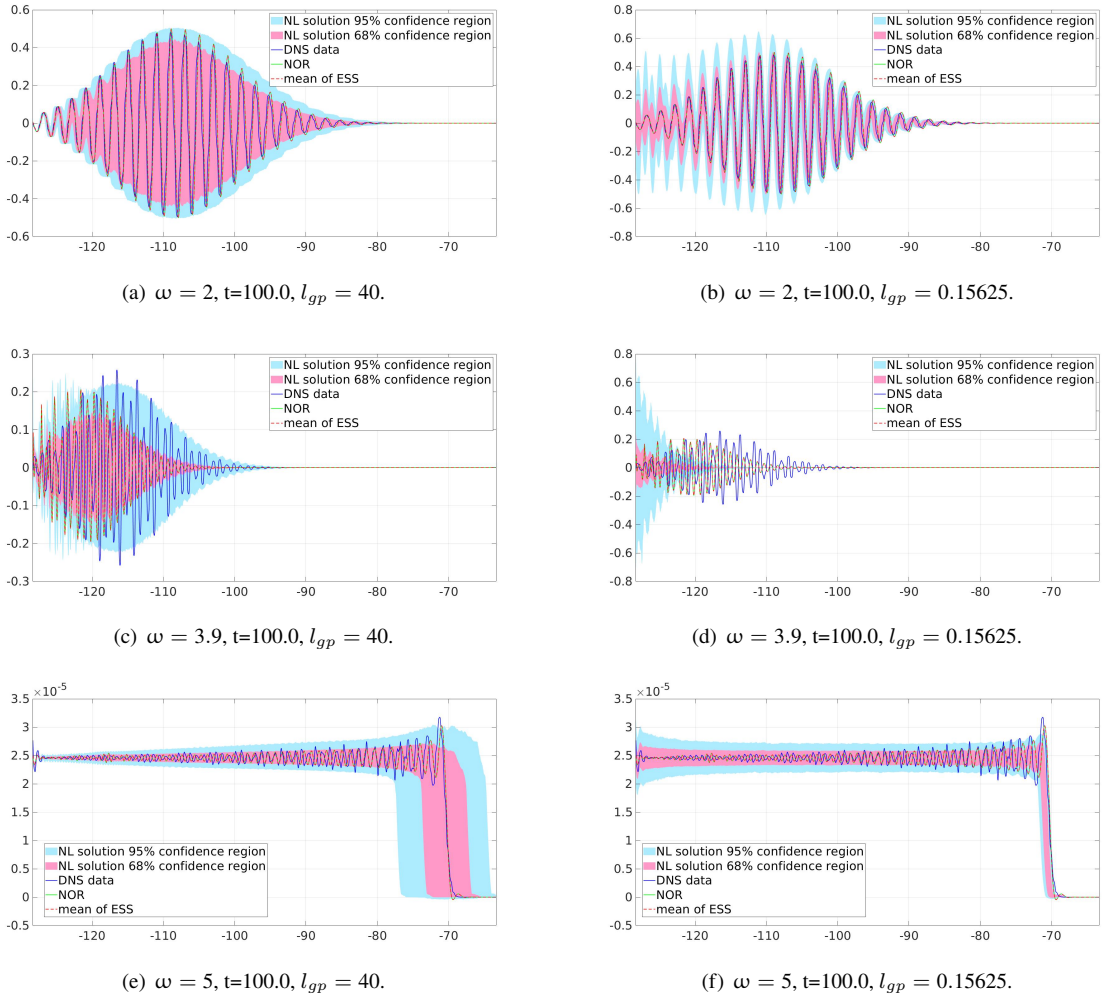
(c) Sample 10 from type 1 data at t=2.0, $l_{gp} = 0.15625$.

(d) Sample 3 from type 2 data at t=2.0, $l_{gp} = 0.15625$.

Fig. 8. Posterior uncertainty on training samples for a bar with periodic microsturcture using two different correlation length.

simulation up to $t = 100$, which is 50$\times$ the training time interval. In Fig.9, we plot the results using $l_{gp} = 40$ and 0.15625. For a low-frequency wave ($\omega = 2.0$), $l_{gp} = 0.15625$ works better, in the sense that the predictive distribution more closely follows the small features of the wave, roughly reflecting the local magnitude of the discrepancy between the two solutions. On the other hand, the case with $l_{gp} = 40$ fails to do so, with a predictive distribution that broadly encompasses the two solutions, but does not capture the small-scale structure. In both cases, the confidence region fully covers the ground truth, providing a conservative estimation of uncertainty, and avoiding overconfidence, as is the intent of the embedded model error construction. Considering next the $\omega = 3.9$ case, a frequency close to the band gap, we find that the larger correlation length $l_{gp} = 40$ provides better predictions. Here, the case with smaller correlation length $l_{gp} = 0.15625$ suffers from a mismatch between the uncertainty prediction, the single push-forward of the mean of ESS, and the ground truth. This observation is consistent with the results in Fig.7, where the band gap shifts from the ground-truth band gap and the confidence region fails to cover the DNS data. For the frequency ($\omega = 5.0$) that is much larger than the band gap frequency, the stress wave is anticipated to stop propagating. In this setting we find that both small and large correlation length cases work well. The best l_{gp} values for the validation samples are also provided in Fig.10, where the optimal correlation length l_{gp} varies depending on the frequency and wave type.

In conclusion, the correlation length for the embedded model error impacts the results differently for different waves, so that the optimal choice would need to be selected according to the purpose of the task. A frequency/waveform-dependent l_{gp} might be of interest. We leave this as a possible future direction. In the following, we choose the best l_{gp} according to the fidelity of prediction of group velocity in Fig.7, choosing $l_{gp} = 2L = 40$ as the optimal correlation length.



(a) $\omega = 2$, t=100.0, $l_{gp} = 40$.

(b) $\omega = 2$, t=100.0, $l_{gp} = 0.15625$.

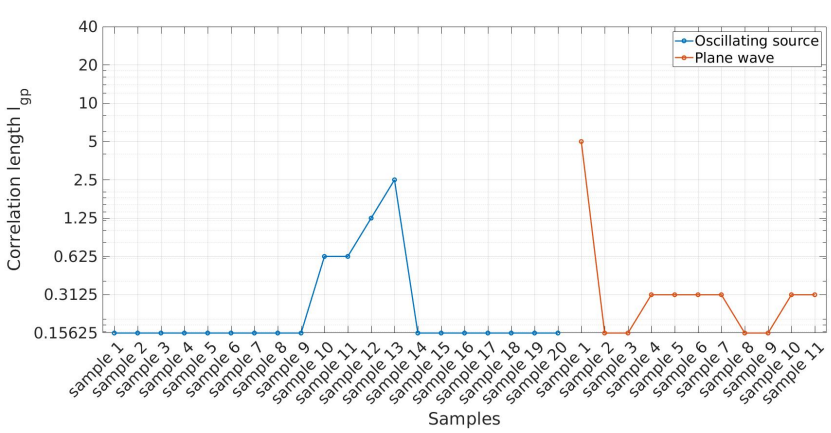
(c) $\omega = 3.9$, t=100.0, $l_{gp} = 40$.

(d) $\omega = 3.9$, t=100.0, $l_{gp} = 0.15625$.

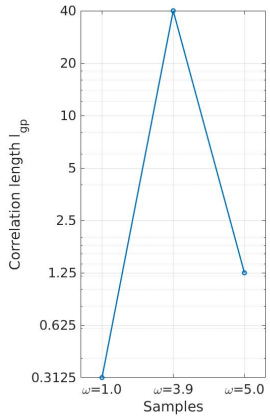
(e) $\omega = 5$, t=100.0, $l_{gp} = 40$.

(f) $\omega = 5$, t=100.0, $l_{gp} = 0.15625$.

Fig. 9. Validation on wave packet for the periodic material at the last time step $t = 100.0$. The columns correspond to different correlation length l_{gp} and the rows correspond to different frequencies ω.



(a) Best l_{gp} for training samples



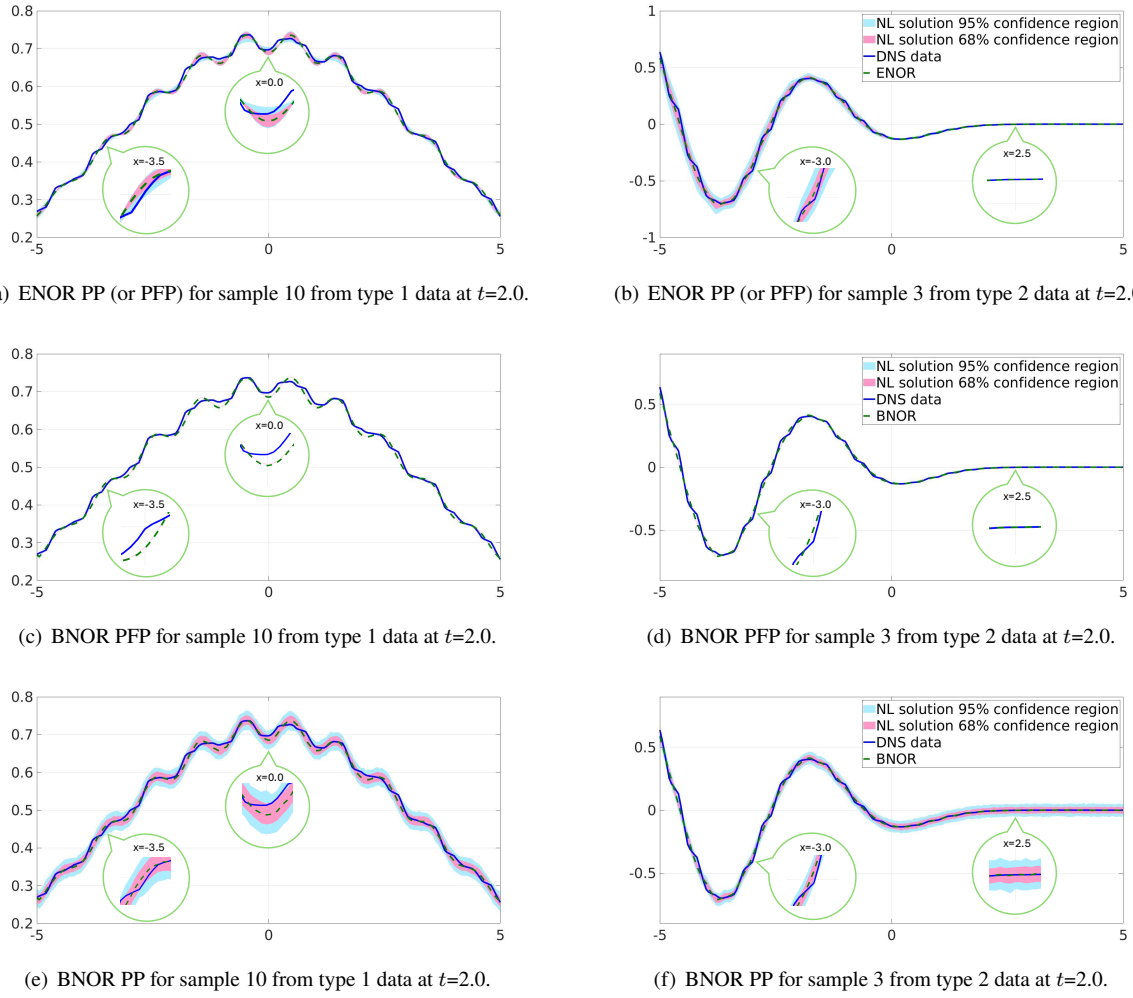
(b) Best l_{gp} for validation samples

Fig. 10. Best l_{gp} (with the lowest CRPS) on training and validation samples for a bar with periodic material.

4.1.3 Comparison with the Baseline

To further illustrate the efficacy of the present ENOR construction in capturing model error, we compare the posterior uncertainty on model predictions (with $l_{gp} = 40$) to the results from BNOR [47], where an additive *iid* noise model was used. Results are shown in Fig.11 for two training data samples at a given time instant $t = 2.0$. We compare the equivalent PP/PFP ENOR results with the corresponding BNOR PP and PFP results. From Figs.11 (c) and (d), one can see that the BNOR PFP exhibits an almost negligible confidence region. On the other hand, the BNOR PP, shown in Figs.11 (e) and (f), exhibits higher uncertainty with a uniform-width confidence region around the mean prediction, as is expected given the additive *iid* noise. Compared with these two baseline results, the uncertainty given by the ENOR PP/PFP, shown in Figs.11 (a) and (b), is somewhat more adaptive, exhibiting a degree of uncertainty that approximately tracks the discrepancy between the mean prediction and the DNS data, highlighting the effectiveness of the embedded model error setting. To provide a quantitative comparison, we note that the BNOR results at $t = 2.0$ in [47] exhibit a PFP CRPS of 0.051 and a PP CRPS of 0.039, which are both roughly $10\times$ higher than the present ENOR result in the worst case (CRPS of 0.0048). To further compare the two models at different regions, we also consider point-wise comparisons at four locations in our exemplar samples. For sample 10 from type 1 data, we consider $x = -3.5$ and $x = 0.0$, which correspond to points in low-uncertainty region and high-uncertainty region, respectively, and highlight these two locations in the left column of Fig.11. At $x = -3.5$, the CRPS for ENOR, BNOR-PFP and BNOR-PP are 0.0031, 0.0046 and 0.0040, respectively. At $x = 0.0$, the CRPS for ENOR, BNOR-PFP and BNOR-PP are 0.0048, 0.0061 and 0.0056. For sample 3 from type 2 data, we choose $x = 2.5$ for investigations in low-uncertainty region

and $x = -2.5$ for high-uncertainty region, and illustrate these in the right column of Fig.11. The CRPS (from top to bottom) are 0.00018, 0.0031, 0.0045 at $x = 2.5$ and 0.028, 0.14, 0.14 at $x = -2.5$. This again illustrates that the ENOR model provides an uncertainty that better reflects the discrepancy between the mean model prediction and the ground truth.



(a) ENOR PP (or PFP) for sample 10 from type 1 data at t=2.0.

(b) ENOR PP (or PFP) for sample 3 from type 2 data at t=2.0.

(c) BNOR PFP for sample 10 from type 1 data at t=2.0.

(d) BNOR PFP for sample 3 from type 2 data at t=2.0.

(e) BNOR PP for sample 10 from type 1 data at t=2.0.

(f) BNOR PP for sample 3 from type 2 data at t=2.0.

Fig. 11. Comparison of posterior uncertainty (PFP and PP) between ENOR and BNOR for periodic material. The columns correspond to different samples in training data and the rows correspond to different methods.

4.1.4 Parametric Uncertainty versus Model Error

It is instructive to examine the role of parametric uncertainty versus model error in the resulting uncertainty in model predictions. We illustrate in Fig.12 the posterior predictive uncertainty under the following three scenarios.

1. In plots (a) and (b), we consider uncertainty from all sources, by sampling realizations from both the embedded error correction term ξ and the marginal distribution on kernel parameters.

2. In plots (c) and (d), we consider the uncertainty from the embedded error correction term ξ only, using deterministic kernel parameters.

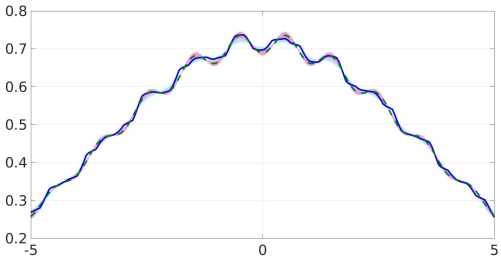
3. In plots (e) and (f), we neglect the embedded error correction term but sample the kernel parameter $\mathbf{C}$ from the learnt marginal distribution. As such, only the uncertainty from the marginal distribution on kernel parameters is considered.

By comparing the confidence regions from these three settings, one can observe that results from scenarios 1 and 2 almost coincide, while the predicted uncertainty in scenario 3 is negligible. This indicates that nearly all of the predictive uncertainty comes from the embedded model error. This is typical in Bayesian estimation of one model against another in the absence of data noise and when there is a sufficiently large amount of data.

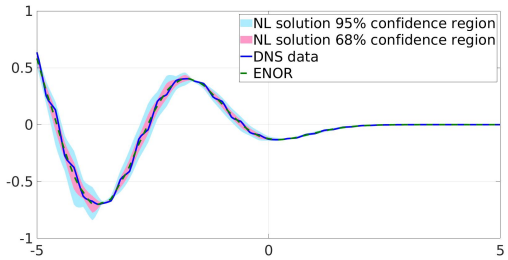
4.2 Example 2: Random Microstructure

To investigate the performance of the ENOR model on more complicated microstructures, we consider here a disordered heterogeneous bar, where the layer lengths of the two materials are random. In particular, the layer sizes are two uniformly distributed random variables: $b_1, b_2 \sim \mathcal{U}[(1-D)b, (1+D)b]$, with the average layer size $b = 0.2$ and the disorder parameter $D = 0.5$. The density for both components is still set as $\rho = 1$ and their Young's moduli are set at $E_1 = 1$ and $E_2 = 0.25$.

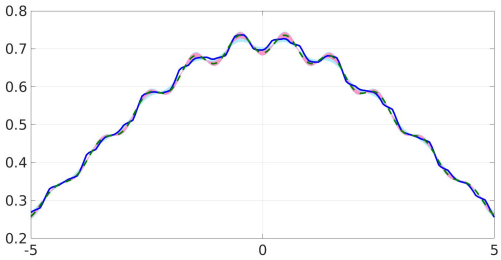
To generate the training dataset, we consider the high-fidelity data under the same settings as in data types 1 and 2 of the periodic bar case. Compared with the periodic microstructure case, from the group velocity generated by the DNS simulations we note that the band stop generally occurs at a lower frequency in the random microstructure case. In fact, for the microstructure considered here, an estimated band stop frequency $\omega_{bs} \approx 3$ can be obtained from the DNS simulations. Therefore, for the validation data set, we study wave packets with frequencies $\omega = 1, 2, 3$ and 4, with the purpose of investigating the performance of our nonlocal surrogate model when the loading frequencies are below ($\omega = 1, 2$), around ($\omega = 3$), and above ($\omega = 4$) the estimated band stop frequency ω_{bs}. We use $\hat{\sigma}^2 = \frac{0.01^2}{0.35}$ and $[\ln(\sigma_{gp}^{lo}), \ln(\sigma_{gp}^{hi})] = [-2.5, -1.7]$ to achieve optimal performance.



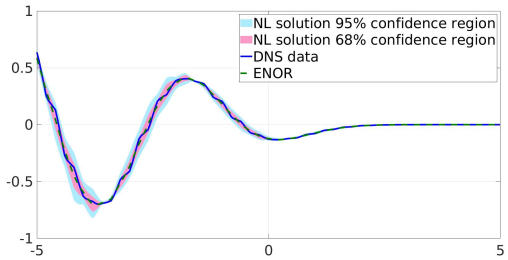
(a) Sample 10 from type 1 data at $t = 2.0$. Confidence region using 100 effective samples and 100 realization each.



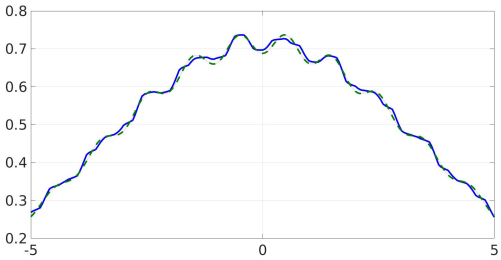
(b) Sample 3 from type 2 data at $t = 2.0$. Confidence region using 100 effective samples and 100 realization each.



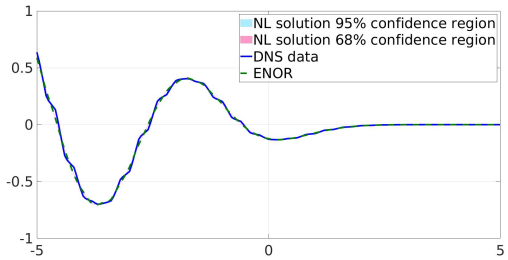
(c) Sample 10 from type 1 data at $t = 2.0$. Confidence region using fixed values of kernel parameters $\mathbf{C}$ and σ_{gp} with 1000 GP realizations.



(d) Sample 3 from type 2 data at $t = 2.0$. Confidence region using fixed values of kernel parameters $\mathbf{C}$ and σ_{gp} with 1000 GP realizations.



(e) Sample 10 from type 1 data at $t = 2.0$. Confidence region using 100 effective samples of $\mathbf{C}$ and σ_{gp} without GP (i.e. with ξ_m=0).



(f) Sample 3 from type 2 data at $t = 2.0$. Confidence region using 100 effective samples of $\mathbf{C}$ and σ_{gp} without GP (i.e. with ξ_m=0).

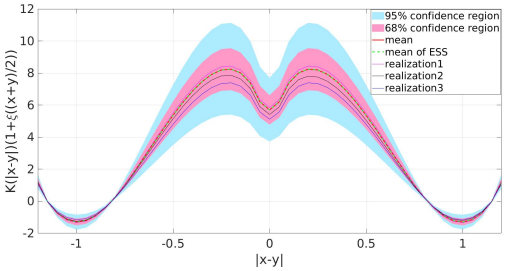
Fig. 12. An illustration of the relative impact of parametric uncertainty and model error on resulting predictive uncertainty, in the periodic material. The columns correspond to different samples in the training data and the rows correspond to posterior prediction using different sources of uncertainty.

4.2.1 Results from MCMC Experiments

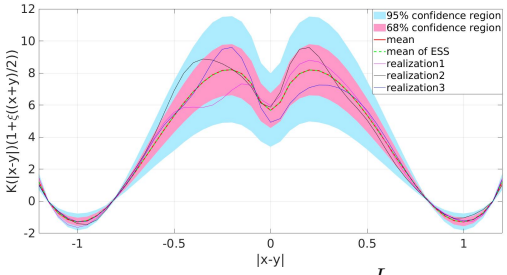
To demonstrate the convergence of MCMC in this case, we again pick correlation length $l_{gp} = L/2$ as an example case to test the utility of ENOR in disordered materials. Following the same settings of the periodic microstructure study, we run 6 independent chains. The results pass the same convergence check both visually and quantitatively, where the improved $\hat{R}$ statistics for all the 24 parameters are very close to 1, with a maximum of 1.0038, which again verifies the convergence. The chains provide 24,000 draws in total, with approximately 39% acceptance rate on average and the combined chain shows good mixing. Compared with the periodic bar with inferred $\ln(\sigma_{gp}) \approx -2.0$, a slightly larger kernel variation, $\ln(\sigma_{gp}) \approx -1.8$, is obtained here. That suggests that the randomness in the material with disordered microstructure results in a larger model discrepancy relative to the DNS results.

4.2.2 Impact of GP Correlation Length

We now study the impact of different correlation length (l_{gp}) values in the disordered material, again considering l_{gp} values ranging from 0.15625 to 40. In Fig.13, we observe that the uncertainty in the kernel with an embedded GP, $K(|x-y|)(1+\xi(\frac{x+y}{2}))$, has the same pattern of mean and confidence region variation as in the periodic bar case, given the same random kernel structure and GP properties. In Fig.14, we investigate the behavior of the group velocity, and observe that a larger correlation length provides better fitting of the group velocity, and that reducing the correlation length l_{gp} results in a left shift of the band gap in prediction. While this trend was also observed in the periodic microstructure case (see Fig.7), severer oscillations and larger confidence regions are observed at low frequencies here in the small correlation length $l_{gp} = 0.625$ case, possibly because of the higher uncertainty levels introduced by the disordered microstructure in the material. In Figs.14 (b) and (d), we show the dispersion curves of the learnt kernels, noting again that positivity of all dispersion curves indicates physical stability of the learnt models.

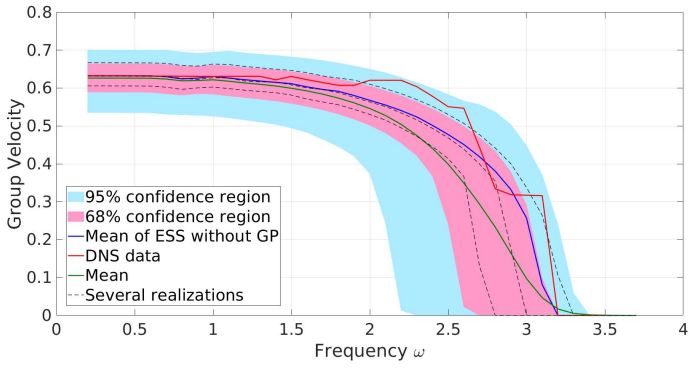


(a) Kernel at point $x = 0.0$ for $l_{gp} = 2L$=40.

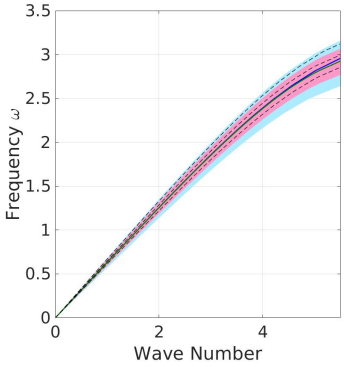


(b) Kernel at point $x = 0.0$ for $l_{gp} = \frac{L}{128}$=0.15625.

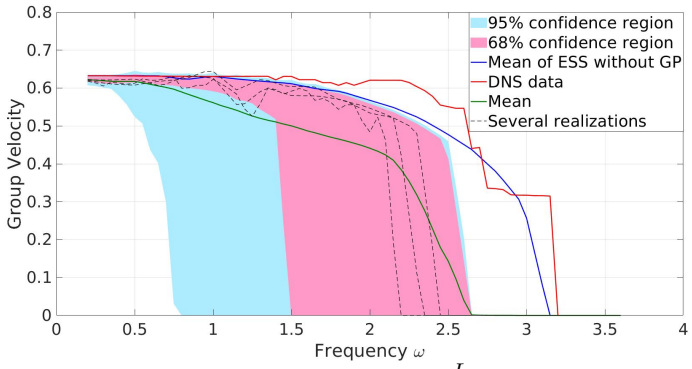
Fig. 13. Kernel with uncertainty for a bar with periodic microstructure using two different correlation lengths.



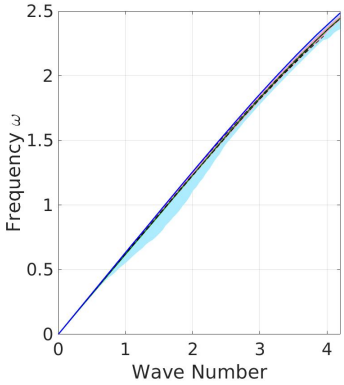
(a) Group velocity using $l_{gp} = 2L$=40.



(b) Dispersion curve using $l_{gp} = 2L$=40.



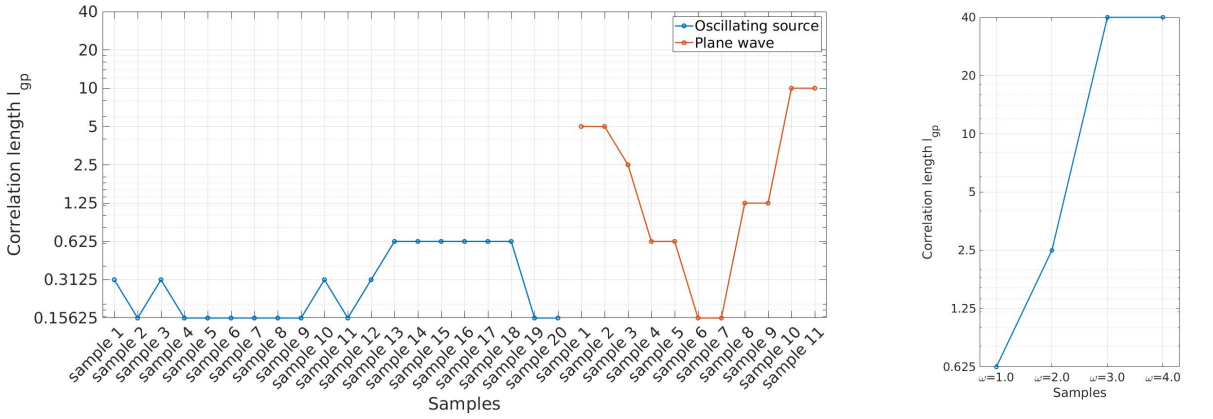
(c) Group velocity using $l_{gp} = \frac{L}{32}$=0.625.



(d) Dispersion curve using $l_{gp} = \frac{L}{32}$=0.625.

Fig. 14. Group velocity and dispersion curve with uncertainty for a bar with disordered microstructure using two different correlation lengths.

Finally, we provide again the prediction of a wave packet for a larger domain and longer time as a validation, not present in the training set. Overall, the same conclusion on the correlation length as section 4.1.2 can be reached for disordered material. We refer the reader to FigB.27 in Appendix B for more details on this experiment. In Fig.15, we provide the best l_{gp} for all the training and validation samples according to the value of the CRPS. Note that the optimal l_{gp}s are different from the ones illustrated in Fig.10, highlighting the dependence of the optimal l_{gp} on the material microstructure.



(a) Best l_{gp} for training samples

(b) Best l_{gp} for validation samples

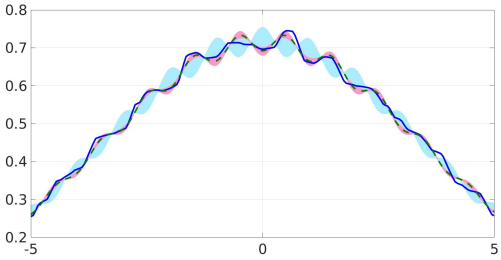
Fig. 15. Best l_{gp} (with the lowest CRPS) on training and validation samples for a bar with disordered microstructure.

4.2.3 Comparison with the Baseline

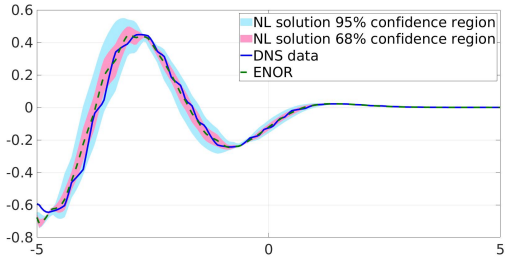
Comparing the CRPS for the posterior PDF on the learnt ENOR model predictions with $l_{gp} = 40$, to that from the baseline model (BNOR) [47], at the same time instant $t = 2.0$, we find again superior performance from the present construction. Specifically, the ENOR CRPS is 0.008 while the BNOR PFP and PP CRPS values are 0.041 and 0.025 respectively. Thus, consistent with the findings in the periodic material, the BNOR CRPS values are much higher than the ENOR CRPS, again indicating that the embedded model error construction provides a better statistical fit for the DNS data.

Further, in Fig.16 we compare the posterior predictive density from ENOR and the BNOR PFP/PP densities for two samples from two wave types at $t = 2$. While ENOR again provides a better prediction for solution uncertainty, the 95% confidence region is significantly larger, especially in Fig.16(a). Note that according to Table.1 and Fig.15, $l_{gp} = 0.3125$ is best for sample 10 in wave type 1, while $l_{gp} = 2.5$ is best for sample 3 in wave type 2. Thus, the correlation length $l_{gp} = 40$ employed here is not ideal for both samples. This again suggests that different correlation

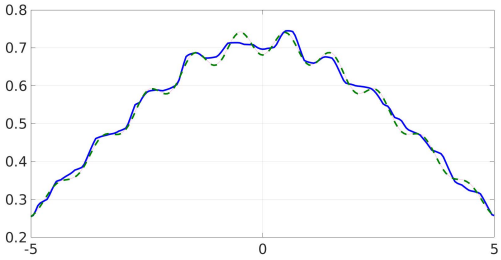
length should ideally be employed for different loading scenarios. When comparing the level of solution uncertainty from periodic (see Fig.11 (a) and (b)) and disordered materials (Fig.16 (a) and (b)), a larger confidence region is observed in the later case. In fact, the average standard deviation over the bar in sample 10 of data setting 1 is 0.0037 in the periodic microstructure case and 0.0122 in the disordered case. Similarly, the average standard deviations in sample 3 are 0.0232 and 0.0262 for the periodic and disordered microstructure cases, respectively. These results highlight the increase in uncertainty when randomness is introduced in the microstructure.



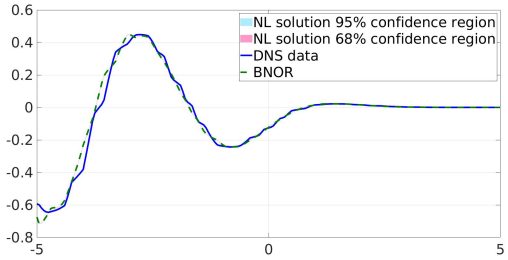
(a) ENOR PP (or PFP) for sample 10 from type 1 data at t=2.0.



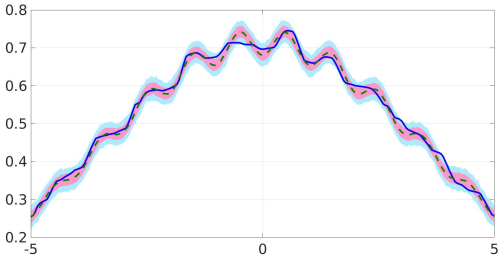
(b) ENOR PP (or PFP) for sample 3 from type 2 data at t=2.0.



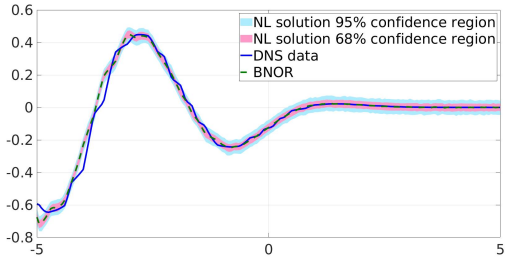
(c) BNOR PFP for sample 10 from type 1 data at t=2.0.



(d) BNOR PFP for sample 3 from type 2 data at t=2.0.



(e) BNOR PP for sample 10 from type 1 data at t=2.0.

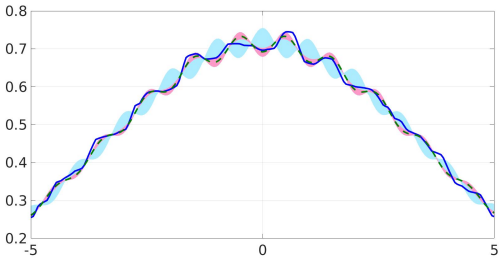


(f) BNOR PP for sample 3 from type 2 data at t=2.0.

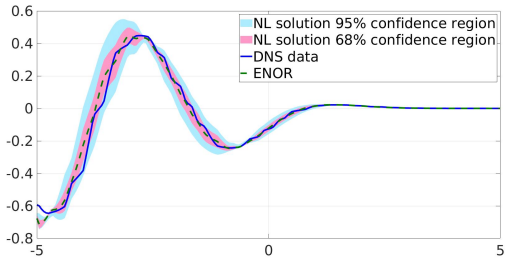
Fig. 16. Comparison of posterior uncertainty (PFP and PP) between ENOR and BNOR for the disordered material. The columns correspond to different samples in training data and the rows correspond to different methods.

4.2.4 Parametric Uncertainty versus Model Error

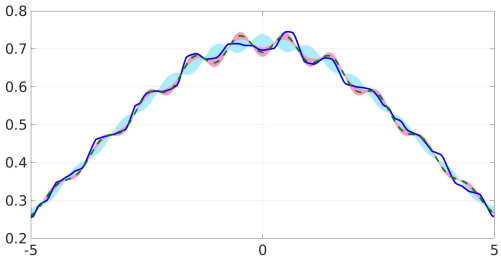
With the purpose of investigating the impact of different source of uncertainty, we examine the predictive uncertainty resulting from parametric uncertainty versus model error for the disordered material, following the same steps in section 4.1.4. As shown in Fig.17, the embedded model error again dominates the posterior uncertainty, with a negligible role for parametric uncertainty.



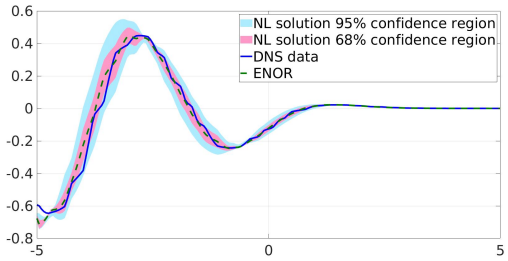
(a) Sample 10 from type 1 data at $t = 2.0$. Confidence region using 100 ess and 100 realization each.



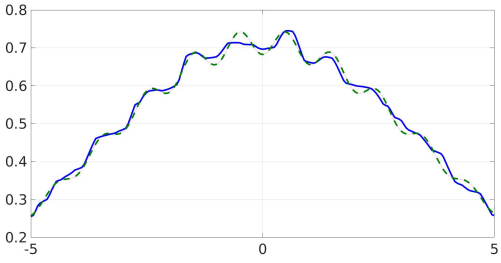
(b) Sample 3 from type 2 data at $t = 2.0$. Confidence region using 100 effective samples and 100 realization each.



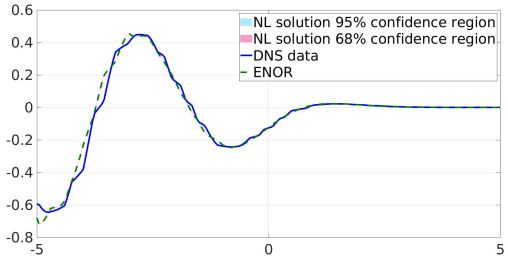
(c) Sample 10 from type 1 data at $t = 2.0$. Confidence region using fixed value of kernel parameters $\mathbf{C}$ and σ_{gp} with 1000 realization of GP.



(d) Sample 3 from type 2 data at $t = 2.0$. Confidence region using fixed value of kernel parameters $\mathbf{C}$ and σ_{gp} with 1000 realization of GP.



(e) Sample 10 from type 1 data at $t = 2.0$. Confidence region using 100 effective samples of kernel parameters $\mathbf{C}$ and σ_{gp} without GP (i.e. set ξ_m=0).



(f) Sample 3 from type 2 data at $t = 2.0$. Confidence region using 100 effective samples of kernel parameters $\mathbf{C}$ and σ_{gp} without GP (i.e. set ξ_m=0).

Fig. 17. An illustration of the relative impact of parametric uncertainty and model error on resulting predictive uncertainty, in the disordered material. The columns correspond to different samples in training data and the rows correspond to posterior prediction using different sources of uncertainty.

5. CONCLUSION

In this work, we proposed ENOR, a novel Bayesian embedded model error framework, to learn the optimal nonlocal surrogate for multiscale homogenization while also characterizing the impact of model error on predictive uncertainty. When learning the bottom-up nonlocal surrogate from microscale simulations, we find that the best fit surrogate has unavoidable discrepancy from the microscale model, rendering a model error UQ study necessary. While the prior work [47] focused on an additive *iid* error construction, we here introduce, for the first time, an embedded model error representation in nonlocal operator learning, capturing the impact of model error on predictive uncertainty. The algorithm is developed by adding a Gaussian process to the nonlocal kernel, such that the kernel parameters and the Gaussian process parameters can be inferred simultaneously. To solve the Bayesian inference problem, a two phase algorithm, Alg.1, and a multilevel delayed acceptance Markov chain Monte Carlo (MLDA-MCMC) method are proposed, to provide efficient sampling and fast converging chains. The effectiveness of ENOR is demonstrated on the stress wave propagation problem in heterogeneous bars. Comparing to the prior work [47], ENOR improves the accuracy in (1) capturing the correct group velocity; (2) producing high-fidelity training data and predicting the substantially different wave type with accurate confidence region structure; (3) posterior sampling of the model parameters; and (4) selection for the correlation length for different tasks.

From both visual inspection and quantitative tests, it was observed that the optimal choice of the GP correlation length may differ for different frequencies and wave types. As a natural extension, we plan to investigate frequency-dependent models. We also plan to explore uncertainty quantification for more complex data-driven homogenization models, such as peridynamics models [22,23] in 2D or 3D, and nonlinear models based on neural networks [34,42, 91]. Since the number of trainable parameters increases substantially in these models, efficient Bayesian inference techniques and reduced order error models would be desired.

ACKNOWLEDGMENTS

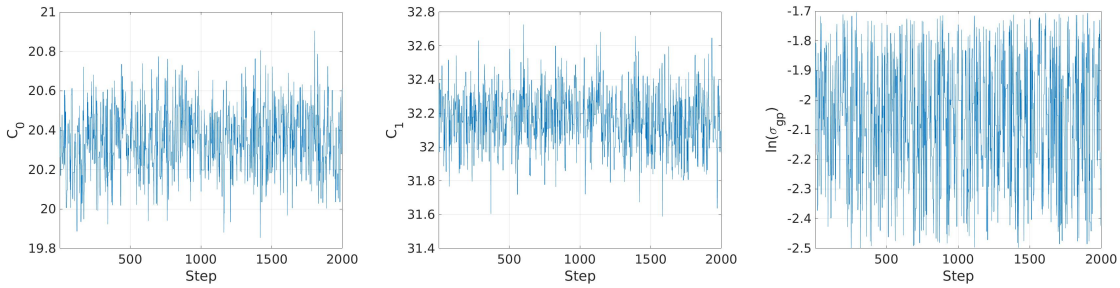
YF and YY acknowledge support by the National Science Foundation under award DMS-2436624, the National Institute of Health under award 1R01GM157589-01, and the AFOSR grant FA9550-22-1-0197. Portions of this research were conducted on Lehigh University's Research Computing infrastructure partially supported by NSF Award 2019035. HNN acknowledges the support of the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research (ASCR), Scientific Discovery through Advanced Computing (SciDAC) Program through the FASTMath Institute. Sandia National Laboratories is a multi-mission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC (NTESS), a wholly owned subsidiary of Honeywell International

Inc., for the U.S. Department of Energy's National Nuclear Security Administration (DOE/NNSA) under contract DE-NA0003525. This written work is authored by an employee of NTESS. The employee, not NTESS, owns the right, title and interest in and to the written work and is responsible for its contents. Any subjective views or opinions that might be expressed in the written work do not necessarily represent the views of the U.S. Government. The publisher acknowledges that the U.S. Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this written work or allow others to do so, for U.S. Government purposes. The DOE will provide public access to results of federally sponsored research in accordance with the DOE Public Access Plan.

APPENDIX A. ABLATION STUDY

Herein, we discuss the tuning procedure of two hyperparameters: the standard deviation of the prior distribution of $\mathbf{C}$, and the lower and upper bounds on σ_{gp}.

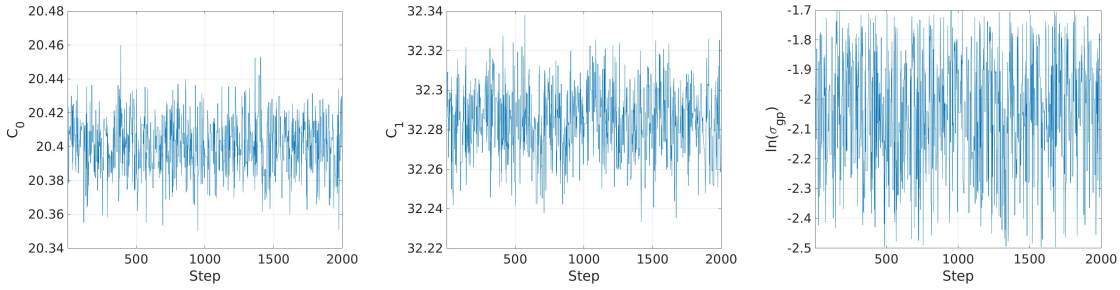
For the tuning of the prior distribution of $\mathbf{C}$ (denoted as $\hat{\sigma}$), we choose the parameter to reach good performance in terms of the ESS and the acceptance rate. In Figures A.18-A.20 we show how $\hat{\sigma}$ affects the MCMC algorithm performances. The results are based on a chain with 2,000 draws with other settings identical to the paper and modifying only $\hat{\sigma}$. Generally, if $\hat{\sigma}$ is too large, the proposed chain samples are more likely to result in low posterior values, resulting in a low acceptance rate (as shown in Fig.A.18). On the other hand, $\hat{\sigma}$ is too small, the resultant strong prior may dominate the likelihood (as shown in Fig.A.20). To achieve a good balance, we choose $\hat{\sigma} = \frac{0.01^2}{0.35}$ for the results in the main text.



(a) First parameter (b) Second parameter (c) $\ln(\sigma_{gp})$

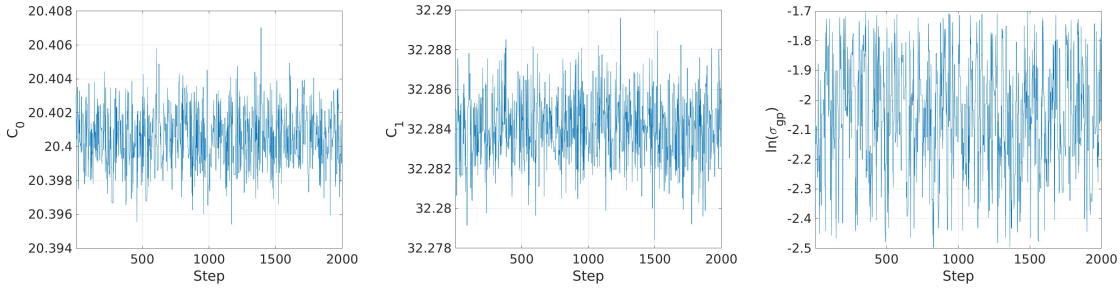
Fig. A.18. Trace plot for the case $\hat{\sigma}^2=\frac{0.1^2}{0.35}$, with acceptance rate ≈ 0.44 and ESS ≈ 685 out of 2,000 draws.

For the tuning of the lower and upper bounds of σ_{gp}, we choose the range to have the Adaptive Error Model (AEM) provide good agreement between the coarse and fine models. Generally, the ABC-likelihood approximated by



(a) First parameter (b) Second parameter (c) $\ln(\sigma_{gp})$

Fig. A.19. Trace plot for the case $\hat{\sigma}^2=\frac{0.01^2}{0.35}$ (the hyperparameter used in this work), with acceptance rate $\approx$ 0.44 and ESS $\approx$ 705 out of 2,000 draws.



(a) First parameter (b) Second parameter (c) $\ln(\sigma_{gp})$

Fig. A.20. Trace plot for the case $\hat{\sigma}^2=\frac{0.001^2}{0.35}$, with acceptance rate $\approx$ 0.44 and ESS $\approx$ 727 out of 2,000 draws. As can be observed, the prior dominates the likelihood and all samples are in a very narrow regime.

the coarse model should not be too far away from that of the fine model. In Fig.A.21 we plot the value of the negative ABC log-likelihood vs the value of $\ln(\sigma_{gp})$. One can see that the two models have small discrepancy when $\ln(\sigma_{gp})$ is between -2.5 and -1.7. To further demonstrate the impact of this range, in Figs. A.22-A.26 we present results from different upper and lower bounds. The results are based on a chain with 2,000 draws. To achieve a high ESS, we choose $[\ln(\sigma_{gp}^{lo}), \ln(\sigma_{gp}^{hi})]$=$[-2.5, -1.7]$ for best performance in the main text.

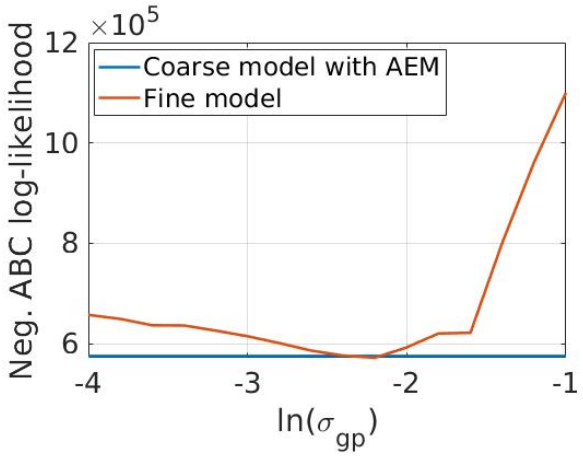
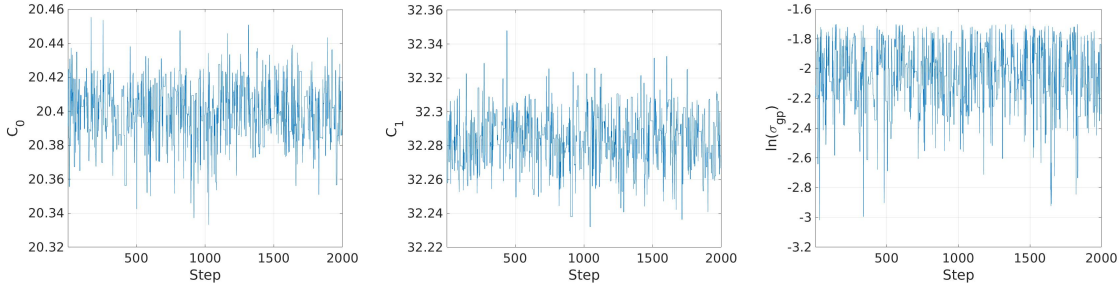


Fig. A.21. Comparison of negative ABC log-likelihood for the fine and coarse (with AEM) model.

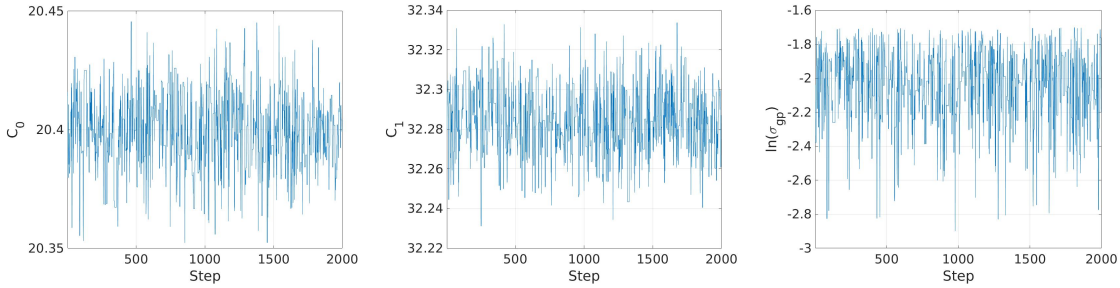


(a) First parameter (b) Second parameter (c) $\ln(\sigma_{gp})$

Fig. A.22. Trace plot for the case $[\ln(\sigma_{gp}^{lo}), \ln(\sigma_{gp}^{hi})]$=$[-3.5, -1.7]$, with acceptance rate $\approx$ 0.37 and ESS $\approx$ 592 out of 2,000 draws.

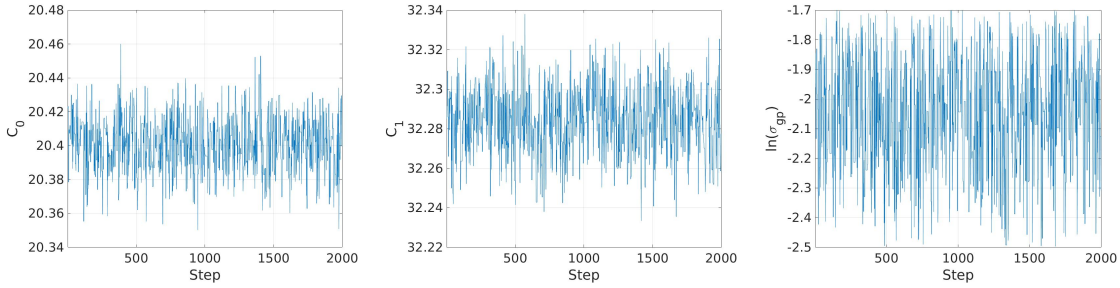
APPENDIX B. ADDITIONAL RESULTS

In this section, we provide additional results for example 2. In Figure B.27, we provide the validation results using $l_{gp} = 40$ and $l_{gp} = 0.15625$ on wave packet for the random microstructure material, at the last time step $t = 100.0$. At low frequencies ($\omega = 1, 2$), $l_{gp} = 0.15625$ works better in reflecting the local magnitude of the solution discrepancy. Near the band stop ($\omega = 3$), $l_{gp} = 40$ works better since the confidence region from the small l_{gp} case fails to cover



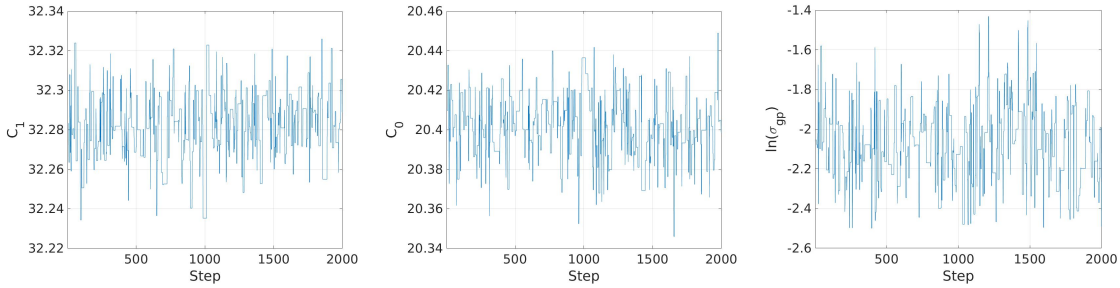
(a) First parameter (b) Second parameter (c) $\ln(\sigma_{gp})$

Fig. A.23. Trace plot for the case $[\ln(\sigma_{gp}^{lo}), \ln(\sigma_{gp}^{hi})]$=$[-3.0, -1.7]$, with acceptance rate $\approx$ 0.38 and ESS $\approx$ 627 out of 2,000 draws.



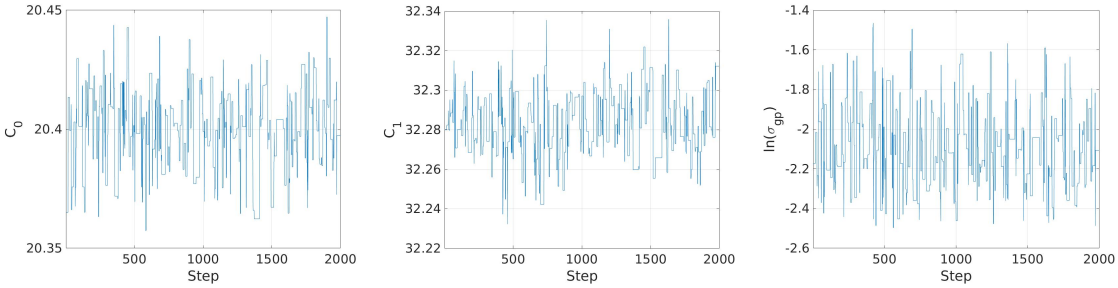
(a) First parameter (b) Second parameter (c) $\ln(\sigma_{gp})$

Fig. A.24. Trace plot for the case $[\ln(\sigma_{gp}^{lo}), \ln(\sigma_{gp}^{hi})]$=$[-2.5, -1.7]$, with acceptance rate $\approx$ 0.44 and ESS $\approx$ 705 out of 2,000 draws.



(a) First parameter (b) Second parameter (c) $\ln(\sigma_{gp})$

Fig. A.25. Trace plot for the case $[\ln(\sigma_{gp}^{lo}), \ln(\sigma_{gp}^{hi})]$=$[-2.5, -1.2]$, with acceptance rate $\approx$ 0.17 and ESS $\approx$ 368 out of 2,000 draws.



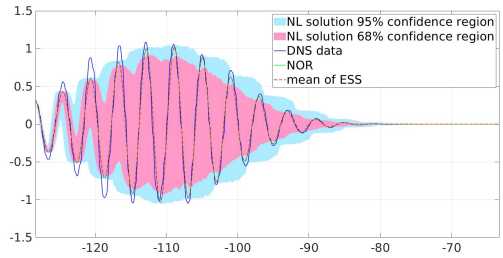
(a) First parameter (b) Second parameter (c) $\ln(\sigma_{gp})$

Fig. A.26. Trace plot for the case $[\ln(\sigma_{gp}^{lo}), \ln(\sigma_{gp}^{hi})]$=$[-3.0, -0.7]$, with acceptance rate $\approx$ 0.15 and ESS $\approx$ 340 out of 2,000 draws.

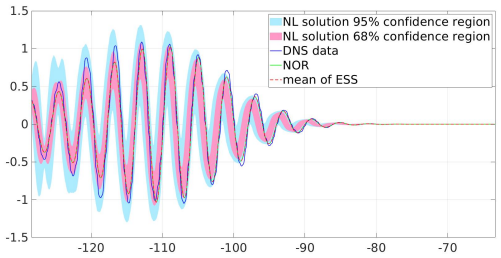
the DNS data. For the large frequency case ($\omega = 4$), both ENOR models have successfully predicted that the stress wave should stop propagating. These observations are consistent with the results from the periodic microstructure material.

REFERENCES

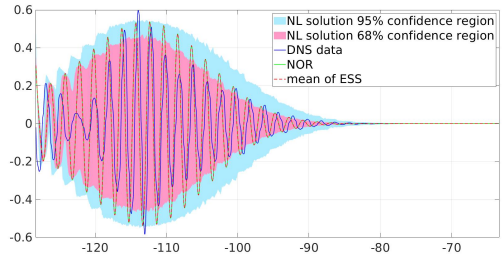
1. Zohdi, T.I., Homogenization methods and multiscale modeling, *Encyclopedia of Computational Mechanics Second Edition*, pp. 1–24, 2017.
2. Bensoussan, A., Lions, J.L., and Papanicolaou, G., *Asymptotic analysis for periodic structures*, Vol. 374, American Mathematical Soc., 2011.
3. Weinan, E. and Engquist, B., Multiscale modeling and computation, *Notices of the AMS*, 50(9):1062–1070, 2003.
4. Efendiev, Y., Galvis, J., and Hou, T.Y., Generalized multiscale finite element methods (gmsfem), *Journal of computational physics*, 251:116–135, 2013.
5. Junghans, C., Praprotnik, M., and Kremer, K., Transport properties controlled by a thermostat: An extended dissipative particle dynamics thermostat, *Soft Matter*, 4(1):156–161, 2008.
6. Kubo, R., The fluctuation-dissipation theorem, *Reports on progress in physics*, 29(1):255, 1966.
7. Santosa, F. and Symes, W.W., A dispersive effective medium for wave propagation in periodic composites, *SIAM Journal on Applied Mathematics*, 51(4):984–1005, 1991.
8. Dobson, M., Luskin, M., and Ortner, C., Sharp stability estimates for the force-based quasicontinuum approximation of homogeneous tensile deformation, *Multiscale Modeling & Simulation*, 8(3):782–802, 2010.
9. Ortiz, M., A method of homogenization of elastic media, *International journal of engineering science*, 25(7):923–934, 1987.



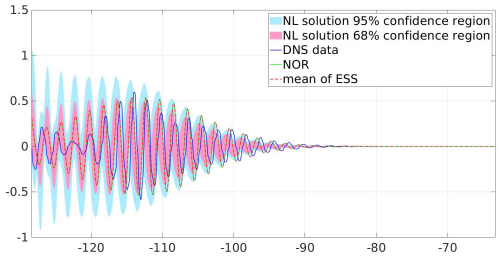
(a) $\omega = 1$, t=100.0, $l_{gp} = 40$.



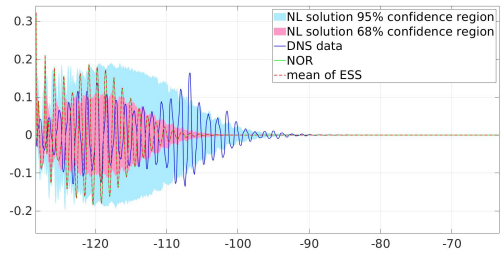
(b) $\omega = 1$, t=100.0, $l_{gp} = 0.15625$.



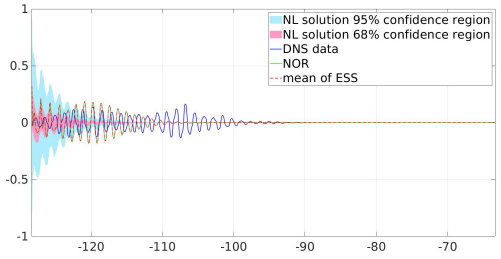
(c) $\omega = 2$, t=100.0, $l_{gp} = 40$.



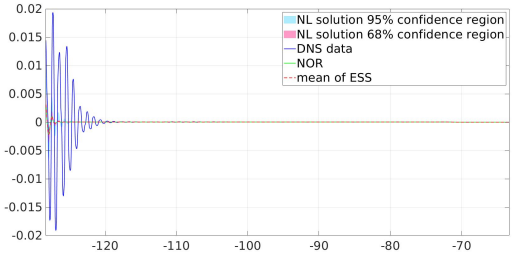
(d) $\omega = 2$, t=100.0, $l_{gp} = 0.15625$.



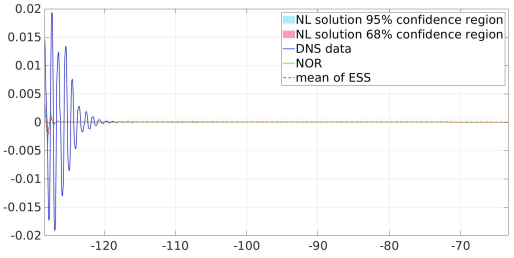
(e) $\omega = 3$, t=100.0, $l_{gp} = 40$.



(f) $\omega = 3$, t=100.0, $l_{gp} = 0.15625$.



(g) $\omega = 4$, t=100.0, $l_{gp} = 40$.



(h) $\omega = 4$, t=100.0, $l_{gp} = 0.15625$.

Fig. B.27. Validation on wave packet for a bar with disordered microstructure at the last time step $t = 100.0$. The columns correspond to different correlation length l_{gp} and the rows correspond to different frequencies ω.

10. Moës, N., Oden, J.T., Vemaganti, K., and Remacle, J.F., Simplified methods and a posteriori error estimation for the homogenization of representative volume elements (rve), *Computer methods in applied mechanics and engineering*, 176(1-4):265–278, 1999.

11. Hughes, T.J., Wells, G.N., and Wray, A.A., Energy transfers and spectral eddy viscosity in large-eddy simulations of homogeneous isotropic turbulence: Comparison of dynamic smagorinsky and multiscale models over a range of discretizations, *Physics of Fluids*, 16(11):4044–4052, 2004.

12. Milton, G.W., *The theory of composites*, SIAM, 2022.

13. Chapman, S.J. and Wilmott, Z.M., Homogenization of flow through periodic networks, *SIAM Journal on Applied Mathematics*, 81(3):1034–1051, 2021.

14. Espanol, P. and Warren, P., Statistical mechanics of dissipative particle dynamics, *Europhysics letters*, 30(4):191, 1995.

15. Groot, R.D. and Warren, P.B., Dissipative particle dynamics: Bridging the gap between atomistic and mesoscopic simulation, *The Journal of chemical physics*, 107(11):4423–4435, 1997.

16. Hoogerbrugge, P. and Koelman, J., Simulating microscopic hydrodynamic phenomena with dissipative particle dynamics, *Europhysics letters*, 19(3):155, 1992.

17. Beran, M. and McCoy, J., Mean field variations in a statistical sample of heterogeneous linearly elastic solids, *International Journal of Solids and Structures*, 6(8):1035–1054, 1970.

18. Cherednichenko, K., Smyshlyaev, V.P., and Zhikov, V., Non-local homogenised limits for composite media with highly anisotropic periodic fibres, *Proceedings of the Royal Society of Edinburgh Section A Mathematics*, 136(1):87–114, 2006.

19. Karal Jr, F.C. and Keller, J.B., Elastic, electromagnetic, and other waves in a random medium, *Journal of Mathematical Physics*, 5(4):537–547, 1964.

20. Rahali, Y., Giorgio, I., Ganghoffer, J., and dell'Isola, F., Homogenization à la piola produces second gradient continuum models for linear pantographic lattices, *International Journal of Engineering Science*, 97:148–172, 2015.

21. Silling, S.A., Propagation of a stress pulse in a heterogeneous elastic bar, *Journal of Peridynamics and Nonlocal Modeling*, 3(3):255–275, 2021.

22. You, H., Yu, Y., Silling, S., and D'Elia, M., A data-driven peridynamic continuum model for upscaling molecular dynamics, *Computer Methods in Applied Mechanics and Engineering*, 389:114400, 2022.

23. You, H., Xu, X., Yu, Y., Silling, S., D'Elia, M., and Foster, J., Towards a unified nonlocal, peridynamics framework for the coarse-graining of molecular dynamics data with fractures, *Applied Mathematics and Mechanics*, 44(7):1125–1150, 2023.

24. You, H., Yu, Y., Silling, S., and D'Elia, M., Nonlocal operator learning for homogenized models: From high-fidelity simulations to constitutive laws, *Journal of Peridynamics and Nonlocal Modeling*, pp. 1–16, 2024.

25. Zhang, L., You, H., and Yu, Y., MetaNOR: A meta-learnt nonlocal operator regression approach for metamaterial modeling, *MRS Communications*, 2022.

26. Tadmor, E. and Tan, C., Critical thresholds in flocking hydrodynamics with non-local alignment, *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 372(2028):20130401, 2014.

27. Du, Q., Engquist, B., and Tian, X., Multiscale modeling, homogenization and nonlocal effects: Mathematical and computational issues, *Contemporary mathematics*, 754, 2020.

28. Silling, S.A., Reformulation of elasticity theory for discontinuities and long-range forces, *Journal of the Mechanics and Physics of Solids*, 48(1):175–209, 2000.

29. Gilboa, G. and Osher, S., Nonlocal operators with applications to image processing, *Multiscale Modeling & Simulation*, 7(3):1005–1028, 2009.

30. Scalas, E., Gorenflo, R., and Mainardi, F., Fractional calculus and continuous time finance, *Physica A*, 284:376–384, 2000.

31. D'Elia, M., Du, Q., Gunzburger, M., and Lehoucq, R., Nonlocal convection-diffusion problems on bounded domains and finite-range jump processes, *Computational Methods in Applied Mathematics*, 29:71–103, 2017.

32. Meerschaert, M. and Sikorskii, A., *Stochastic models for fractional calculus*, Studies in mathematics, Gruyter, 2012.

33. Silling, S.A., Jafarzadeh, S., and Yu, Y., Peridynamic models for random media found by coarse graining, *Journal of Peridynamics and Nonlocal Modeling*, pp. 1–30, 2024.

34. Jafarzadeh, S., Silling, S., Zhang, L., Ross, C., Lee, C.H., Rahman, S., Wang, S., and Yu, Y., Heterogeneous peridynamic neural operators: Discover biotissue constitutive law and microstructure from digital image correlation measurements, *arXiv preprint arXiv:2403.18597*, 2024.

35. Smyshlyaev, V.P. and Cherednichenko, K.D., On rigorous derivation of strain gradient effects in the overall behaviour of periodic heterogeneous media, *Journal of the Mechanics and Physics of Solids*, 48(6-7):1325–1357, 2000.

36. Willis, J.R., The nonlocal influence of density variations in a composite, *International Journal of Solids and Structures*, 21(7):805–817, 1985.

37. Eringen, A.C. and Edelen, D.G.B., On nonlocal elasticity, *International Journal of Engineering Science*, 10(3):233–248, 1972.

38. Bobaru, F., Foster, J.T., Geubelle, P.H., and Silling, S.A., *Handbook of peridynamic modeling*, CRC press, 2016.

39. You, H., Yu, Y., Silling, S., and D'Elia, M., Data-driven learning of nonlocal models: from high-fidelity simulations to constitutive laws, *AAAI Spring Symposium: MLPS*, 2021.

40. You, H., Yu, Y., Trask, N., Gulian, M., and D'Elia, M., Data-driven learning of nonlocal physics from high-fidelity synthetic data, *Computer Methods in Applied Mechanics and Engineering*, 374:113553, 2021.

41. Xu, X., D'Elia, M., and Foster, J.T., A machine-learning framework for peridynamic material models with physical constraints,

Computer Methods in Applied Mechanics and Engineering, 386:114062, 2021.

42. Jafarzadeh, S., Silling, S., Liu, N., Zhang, Z., and Yu, Y., Peridynamic neural operators: A data-driven nonlocal constitutive model for complex material responses, *arXiv preprint arXiv:2401.06070*, 2024.
43. de Moraes, E.A.B., D'Elia, M., and Zayernouri, M., Machine learning of nonlocal micro-structural defect evolutions in crystalline materials, *Computer Methods in Applied Mechanics and Engineering*, 403:115743, 2023.
44. Xu, X., D'Elia, M., Glusa, C., and Foster, J.T., Machine-learning of nonlocal kernels for anomalous subsurface transport from breakthrough curves, *arXiv preprint arXiv:2201.11146*, 2022.
45. Yu, Y., Liu, N., Lu, F., Gao, T., Jafarzadeh, S., and Silling, S., Nonlocal attention operator: Materializing hidden knowledge towards interpretable physics discovery, *arXiv preprint arXiv:2408.07307*, 2024.
46. Lu, F., An, Q., and Yu, Y., Nonparametric learning of kernels in nonlocal operators, *arXiv preprint arXiv:2205.11006*, 2022.
47. Fan, Y., D'Elia, M., Yu, Y., Najm, H.N., and Silling, S., Bayesian nonlocal operator regression: A data-driven learning framework of nonlocal models with uncertainty quantification, *Journal of Engineering Mechanics*, 149(8):04023049, 2023.
48. Laplace, M.d., *Essai philosophique sur les probabilités*, Gauthier-Villars, Paris, 1814. English Ed.: "A Philosophical Essay on Probabilities", Dover Pub., 6^{th} ed., 1995.
49. Casella, G. and Berger, R.L., *Statistical inference*, Vol. 70, Duxbury Press Belmont, CA, 1990.
50. Jaynes, E., *Probability Theory: The Logic of Science,* G.L. Bretthorst, Ed, Cambridge University Press, Cambridge, UK, 2003.
51. Robert, C. and Casella, G., *Monte Carlo Statistical Methods*, Springer Texts in Statistics, Springer, 2004.
52. Sivia, D.S. and Skilling, J., *Data Analysis: A Bayesian Tutorial, Second Edition*, Oxford University Press, 2006.
53. Carlin, B.P. and Louis, T.A., *Bayesian Methods for Data Analysis*, Chapman and Hall/CRC, Boca Raton, FL, 2011.
54. Kennedy, M.C. and O'Hagan, A., Bayesian calibration of computer models, *Journal of the Royal Statistical Society: Series B*, 63(3):425–464, 2001.
55. Higdon, D., Lee, H., and Holloman, C., Markov chain Monte Carlo-based approaches for inference in computationally intensive inverse problems, *Bayesian Statistics*, 7:181–197, 2003.
56. Oliver, T.A. and Moser, R.D., Bayesian uncertainty quantification applied to RANS turbulence models, *J. Phys.: Conf. Ser.*, 318, 2011.
57. Cui, T., Marzouk, Y., and Willcox, K., Scalable posterior approximations for large-scale Bayesian inverse problems via likelihood-informed parameter and state reduction, *Journal of Computational Physics*, 315:363–387, 2016.
58. Hakim, L., Lacaze, G., Khalil, M., Sargsyan, K., Najm, H., and Oefelein, J., Probabilistic parameter estimation in a 2-step chemical kinetics model for n-dodecane jet autoignition, *Combustion Theory and Modelling*, 22(3):446–466, 2018.

59. Huan, X., Safta, C., Sargsyan, K., Geraci, G., Eldred, M.S., Vane, Z.P., Lacaze, G., Oefelein, J.C., and Najm, H.N., Global Sensitivity Analysis and Estimation of Model Error, toward Uncertainty Quantification in Scramjet Computations, *AIAA Journal*, 56(3):1170–1184, 2018.

60. Bayarri, M.J., Berger, J.O., Paulo, R., Sacks, J., Cafeo, J.A., Cavendish, J., Lin, C.H., and Tu, J., A framework for validation of computer models, *Technometrics*, 49(2):138–154, 2007.

61. Kleijnen, J.P., Kriging metamodeling in simulation: A review, *European journal of operational research*, 192(3):707–716, 2009.

62. Sargsyan, K., Najm, H., and Ghanem, R., On the statistical calibration of physical models, *International Journal of Chemical Kinetics*, 47(4):246–276, 2015.

63. Sargsyan, K., Huan, X., and Najm, H.N., Embedded model error representation for bayesian model calibration, *International Journal for Uncertainty Quantification*, 9(4), 2019.

64. Lykkegaard, M.B., Mingas, G., Scheichl, R., Fox, C., and Dodwell, T.J. Multilevel delayed acceptance mcmc with an adaptive error model in pymc3, 2020. Workshop on machine learning for engineering modeling, simulation and design, NeurIPS 2020.

65. Lykkegaard, M.B., Dodwell, T.J., Fox, C., Mingas, G., and Scheichl, R., Multilevel delayed acceptance mcmc, *SIAM/ASA Journal on Uncertainty Quantification*, 11(1):1–30, 2023.

66. Liu, N., Yu, Y., You, H., and Tatikola, N., Ino: Invariant neural operators for learning complex physical systems with momentum conservation, In *International Conference on Artificial Intelligence and Statistics*, pp. 6822–6838. PMLR, 2023.

67. Liu, N., Fan, Y., Zeng, X., Klöwer, M., ZHANG, L., and Yu, Y., Harnessing the power of neural operators with automatically encoded conservation laws, In *Forty-first International Conference on Machine Learning*, 2024.

68. Du, Q., Tao, Y., and Tian, X., A peridynamic model of fracture mechanics with bond-breaking, *Journal of Elasticity*, pp. 1–22, 2017.

69. Kennedy, M.C. and O'Hagan, A., Bayesian calibration of computer models (with discussion), *Journal of the Royal Statistical Society B*, 63:425–464, 2001.

70. Kennedy, M.C., O'Hagan, A., and Higgins, N., Bayesian analysis of computer code outputs, In *Quantitative Methods for Current Environmental Issues. C. W. Anderson, V. Barnett, P. C. Chatwin, and A. H. El-Shaarawi (eds.)*, pp. 227–243. Springer-Verlag: London, 2002.

71. Higdon, D., Kennedy, M.C., Cavendish, J., Cafeo, J., and Ryne, R.D., Combining field data and computer simulations for calibration and prediction, *SIAM Journal on Scientific Computing*, 26:448–466, 2004.

72. Bayarri, M.J., Berger, J.O., Kennedy, M.C., Kottas, A., Paulo, R., Sacks, J., Cafeo, J.A., Lin, C.H., and Tu, J., Predicting vehicle crashworthiness: validation of computer models for functional and hierarchical data, *Journal of the American Statistical Association*, 104:929–942, 2009.

73. Hakim, L., Lacaze, G., Khalil, M., Sargsyan, K., Najm, H., and Oefelein, J., Probabilistic parameter estimation in a 2-step chemical kinetics model for n-dodecane jet autoignition, *Combustion Theory and Modelling*, 22(3):446–466, 2018.

74. Huan, X., Safta, C., Sargsyan, K., Geraci, G., Eldred, M.S., Vane, Z.P., Lacaze, G., Oefelein, J.C., and Najm, H.N., Global sensitivity analysis and estimation of model error, toward uncertainty quantification in scramjet computations, *AIAA Journal*, 56(3):1170–1184, 2018.

75. Hegde, A., Weiss, E., Windl, W., Najm, H.N., and Safta, C., A bayesian calibration framework with embedded model error for model diagnostics, *International Journal for Uncertainty Quantification*, 14(6):37–70, 2024.

76. Beaumont, M., Zhang, W., and Balding, D.J., Approximate Bayesian Computation in population genetics, *Genetics*, 162(4):2025–2035, 2002.

77. Marjoram, P., Molitor, J., Plagnol, V., and Tavaré, S., Markov chain Monte Carlo without likelihoods, *Proc Natl Acad Sci USA*, 100(26):15324–15328, 2003.

78. Sisson, S.A. and Fan, Y. *Handbook of Markov Chain Monte Carlo*, chapter Likelihood-free Markov chain Monte Carlo, pp. 313–338. Chapman and Hall/CRC Press, 2011.

79. Karhunen, K., Zur spektraltheorie stochastischer prozesse, *Ann. Acad. Sci. Fennicae*, 37, 1946.

80. Spanos, P.D., Beer, M., and Red-Horse, J., Karhunen–loève expansion of stochastic processes with a modified exponential covariance kernel, *Journal of Engineering Mechanics*, 133(7):773–779, 2007.

81. Lucor, D., Su, C.H., and Karniadakis, G.E., Generalized polynomial chaos and random oscillators, *International Journal for Numerical Methods in Engineering*, 60(3):571–596, 2004.

82. Cui, T., Fox, C., and O'sullivan, M., Bayesian calibration of a large-scale geothermal reservoir model by a new adaptive delayed acceptance metropolis hastings algorithm, *Water Resources Research*, 47(10), 2011.

83. Abrate, S., Wave propagation in lightweight composite armor, In *Journal de Physique IV (Proceedings)*, Vol. 110, pp. 657–662. EDP sciences, 2003.

84. Abril-Pla, O., Andreani, V., Carroll, C., Dong, L., Fonnesbeck, C.J., Kochurov, M., Kumar, R., Lao, J., Luhmann, C.C., Martin, O.A., , Pymc: a modern, and comprehensive probabilistic programming framework in python, *PeerJ Computer Science*, 9:e1516, 2023.

85. Ter Braak, C.J. and Vrugt, J.A., Differential evolution markov chain with snooker updater and fewer chains, *Statistics and Computing*, 18:435–446, 2008.

86. Vehtari, A., Gelman, A., Simpson, D., Carpenter, B., and Bürkner, P.C., Rank-normalization, folding, and localization: An improved $\hat{R}$ for assessing convergence of mcmc (with discussion), *Bayesian analysis*, 16(2):667–718, 2021.

87. Vats, D., Flegal, J.M., and Jones, G.L., Multivariate output analysis for markov chain monte carlo, *Biometrika*, 106(2):321–337, 2019.

88. Gelman, A., Carlin, J., Stern, H., and Rubin, D., *Bayesian Data Analysis*, Chapman and Hall/CRC, London, 1995.

89. Gelman, A., Meng, X.L., and Stern, H., Posterior predictive assessment of model fitness via realized discrepancies, *Statistica Sinica*, 6(4):733–760, 1996.

90. Zamo, M. and Naveau, P., Estimation of the continuous ranked probability score with limited information and applications to ensemble weather forecasts, *Mathematical Geosciences*, 50(2):209–234, 2018.

91. You, H., Zhang, Q., Ross, C.J., Lee, C.H., and Yu, Y., Learning deep implicit fourier neural operators (ifnos) with applications to heterogeneous material modeling, *arXiv preprint arXiv:2203.08205*, 2022.