

Bayesian Nonlocal Operator Regression (BNOR): A Data-Driven Learning Framework of Nonlocal Models with Uncertainty Quantification

Yiming Fan¹, Marta D’Elia², Yue Yu³, Habib N. Najm⁴, and Stewart Silling⁵

¹Department of Mathematics, Lehigh University, Bethlehem, PA 18015, USA. Email: yif319@lehigh.edu

²Sandia National Laboratories, Livermore, CA, USA. Email: marti.delia@gmail.com

³Corresponding author. Department of Mathematics, Lehigh University, Bethlehem, PA 18015, USA. Email: yuy214@lehigh.edu

⁴Sandia National Laboratories, Livermore, CA, USA. Email: hnnajm@sandia.gov

⁵Center for Computing Research, Sandia National Laboratories, Albuquerque, NM, USA. Email: sasilli@sandia.gov

ABSTRACT

We consider the problem of modeling heterogeneous materials where micro-scale dynamics and interactions affect global behavior. In the presence of heterogeneities in material microstructure it is often impractical, if not impossible, to provide quantitative characterization of material response. The goal of this work is to develop a Bayesian framework for uncertainty quantification (UQ) in material response prediction when using nonlocal models. Our approach combines the nonlocal operator regression (NOR) technique and Bayesian inference. Specifically, [additive independent identically distributed Gaussian noise is employed to model the discrepancy between the nonlocal model and the data](#). Then, we use a Markov chain Monte Carlo (MCMC) method to sample the posterior probability distribution on parameters involved in the nonlocal constitutive law, and associated modeling discrepancies relative to higher fidelity computations. [As an application, we consider the propagation of stress waves through a one-dimensional heterogeneous bar with](#)

randomly generated microstructure. Several numerical tests illustrate the construction, enabling UQ in nonlocal model predictions. Although nonlocal models have become popular means for homogenization, their statistical calibration with respect to high-fidelity models has not been presented before. This work is a first step in this direction, focused on Bayesian parameter calibration.

INTRODUCTION

Nonlocal models have become viable alternatives to partial differential equation (PDE) models when the effects of the small-scale behavior of a system affect its global state (Beran and McCoy 1970; Cherednichenko et al. 2006; Karal Jr and Keller 1964; Rahali et al. 2015; Smyshlyaev and Cherednichenko 2000; Willis 1985; Eringen and Edelen 1972; Bobaru et al. 2016; Du et al. 2020). These models are characterized by integral operators (as opposed to differentiable operators) that embed time and length scales in their definition; as such, they are able to capture long-range effects that classical PDE models fail to describe (Silling 2000). These effects include the anomalous behavior often observed in coarse-grained measurements of a diffusive quantity, often referred to as anomalous or non-standard diffusion (Du and Zhou 2011; Du et al. 2013; Suzuki et al. 2022). In the context of materials response, it is often impractical to solve equations at the small (e.g. micro) scale either because small-scale properties are not available and/or uncertain, or because the simulations are computationally infeasible. Thus, a lot of effort has been dedicated to homogenization theory with the purpose of designing large-scale models that accurately reproduce the effects of small-scale behavior (Zohdi 2017; Bensoussan et al. 2011; Weinan and Engquist 2003; Efendiev et al. 2013; Junghans et al. 2008; Kubo 1966; Santosa and Symes 1991; Dobson et al. 2010; Ortiz 1987; Moës et al. 1999; Hughes et al. 2004). A coarse-grained model may or may not use the same governing equations as the underlying small-scale model. Classical homogenization techniques, such as the use of effective properties to represent the constitutive properties of a heterogeneous medium, have achieved many successes and wide usage. However, they do not always lend themselves to applications in which the size scale of the heterogeneity affects the global response. An example is the propagation of high frequency stress waves in a composite material. As shown in Figure 1, the

experimentally measured wave velocity decreases with frequency (Tauchert and Guzelsu 1972). However, a classical model that uses the effective elastic modulus fails to reproduce this effect. Thus, a nonlocal approach is more compatible with this application than classical homogenization.

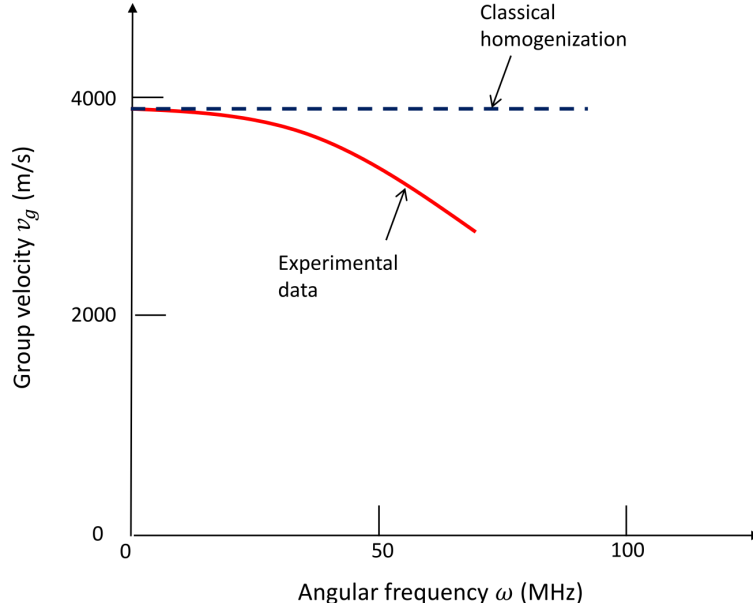


Fig. 1. Dispersion curves for stress waves in a fiber-reinforced composite. Solid line: experiment (redrawn from (Tauchert and Guzelsu 1972)). Dashed line: classical homogenization using effective properties.

Due to their ability to capture internal length scales and long-range effects, nonlocal models are the best candidates for this type of application. (Du et al. 2020). Their use in the context of homogenization is not new and started with the development of homogenized models for subsurface transport (Suzuki et al. 2022). Here, fractional models were identified as the best means for describing super- and sub-diffusive effects, and a Bayesian approach was developed in (Trillos and Sanz-Alonso 2017), to recover the order and the diffusion coefficient of an elliptic fractional partial differential equation. More recently, the concept of nonlocal homogenization was successfully applied in the context of material response (Silling 2021; Silling et al. 2022). Furthermore, with the explosion of machine learning, optimized nonlocal models were designed with the purpose of accurately reproducing observed coarse-grained behavior and predicting unseen behavior with the

learnt model. We refer the reader to (You et al. 2021a; You et al. 2021b; You et al. 2022; Xu et al. 2021; Xu et al. 2022; Zhang et al. 2022; de Moraes et al. 2022) for several examples of the use of machine learning for the design of homogenized nonlocal operators and the rigorous analysis of its learning theory (Lu et al. 2022; Zhang et al. 2022). Although successful in providing a deterministic answer to the homogenization problem, none of these approaches characterize the discrepancy between the homogenized nonlocal surrogate model and the high-fidelity microscale data and its associated prediction uncertainty.

The characterization of surrogate/reduced-order modeling discrepancy is a well-known problem in the PDE literature, while it is absent in the context of nonlocal modeling. In this work, for the first time, we employ Bayesian inference for statistical estimation of nonlocal model parameters using statistical descriptions for the discrepancy from high-fidelity data. Bayesian methods (Laplace 1814; Casella and Berger 1990; Jaynes 2003; Robert and Casella 2004; Sivia and Skilling 2006; Carlin and Louis 2011) have been used increasingly over recent decades for the probabilistic estimation of parameters of complex physical models given data on model predictions (Kennedy and O’Hagan 2001; Higdon et al. 2003; Oliver and Moser 2011; Cui et al. 2016; Hakim et al. 2018; Huan et al. 2018). The Bayesian framework provides a number of advantages in this regard, including means of (1) incorporating prior information which can be based on expert knowledge, prior/other experiments, or physical laws; (2) using arbitrary data models that can incorporate physical model complexity as well as generalized statistical models of discrepancy between data and model predictions; and (3) model comparison, selection, and averaging, with formal connections to information theory. We use Bayesian inference here, relying on an adaptive Markov chain Monte Carlo (MCMC) construction (Haario et al. 2001), to estimate parameters of a homogenized nonlocal model given data from a higher-fidelity PDE model. In this context, the discrepancy between model predictions and the data is due to predictive errors resulting from the necessary approximations employed in the homogenized nonlocal model construction. While more sophisticated statistical models with nontrivial correlation structure are available for the representation of this model error discrepancy (Kennedy and O’Hagan 2000; Sargsyan et al. 2015; Sargsyan et al. 2019), our present

construction is at more basic level, where we use additive independent identically distributed (*i.i.d.*) noise to model this discrepancy.

Specifically, we employ Bayesian calibration in the context of a one-dimensional, nonlocal wave equation that describes the propagation of stress waves through an elastic bar with a heterogeneous microstructure. We use Bayesian inference to learn the posterior probability distributions of parameters of the nonlocal constitutive law (i.e. the nonlocal kernel function). The data model includes an additive noise term to represent the discrepancy between the high-fidelity, PDE models and the (homogenized) nonlocal model. Our key contributions are summarized below.

1. We introduce and demonstrate, for the first time, a Bayesian inference framework for the statistical estimation of parameters in homogenized nonlocal models.
2. We develop a multi-step algorithm that enables efficient Bayesian inference via effective initialization and hyper-parameter choice.
3. Through several numerical tests we illustrate the prediction capability of the proposed approach by successfully using the calibrated model to predict unseen scenarios.

Paper outline. In Section “Dispersion in Heterogeneous Materials” we introduce the classical problem of stress wave propagation through heterogeneous materials and describe the numerical solver used for the generation of the high-fidelity data set. In Section “Background and Related Mathematical Formulation” we summarize relevant prior work on (deterministic) NOR: we describe the mathematical formulation of the regression problem and discuss the choice of hyperparameters for efficient training. In Section “Bayesian Nonlocal Operator Regression” we introduce a Bayesian approach to nonlocal model calibration. We break down the MCMC-based algorithm in several steps with the purpose of initializing distributions and hyperparameters in an efficient manner. In Section “Application to a Heterogeneous Elastic Bar” we illustrate through several numerical tests how our approach predicts the posterior distribution of the nonlocal model predictions, enabling uncertainty quantification (UQ) in nonlocal simulations. Section “Conclusion” concludes the paper with a summary of our contributions and a list of potential follow-up work.

DISPERSION IN HETEROGENEOUS MATERIALS

The classical (local) wave equation for a heterogeneous one-dimensional linearly elastic body reads as follows:

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

where x is the reference position, ρ is the mass density, t is time, u is the displacement, f is the external load density (assumed to be continuous), and E is the elastic modulus which varies spatially. The quantity in the large parentheses in (1) is the stress, usually denoted by σ , which is continuous even if E has jump discontinuities. If E has a jump discontinuity at some x_0 , the following jump conditions hold:

$$[u] = 0, \quad \left[E \frac{\partial u}{\partial x} \right] = 0 \quad (2)$$

where the notation for jumps is $[w] = w(x_0^+) - w(x_0^-)$ for any function w . In this paper, E is piecewise continuous. The particular choice of the function $E(x)$ describes a microstructure. For simplicity, E is assumed to take on only one of two values, but the spatial distribution can be either periodic or random (see the left plot of Figure 2). Also for simplicity, ρ is assumed to be constant throughout the body. Open subregions that have constant E will be called “grains” in analogy with materials science.

In the case of a random microstructure, the grain sizes L_1 and L_2 for materials 1 and 2 are generated using the following expressions:

$$\begin{aligned} L_1 &\sim \mathcal{U}[(1-D)(1-\phi)L, (1+D)(1-\phi)L] \\ L_2 &\sim \mathcal{U}[(1-D)(1+\phi)L, (1+D)(1+\phi)L] \end{aligned} \quad (3)$$

where L is the mean value of grain size and ϕ determines the volume fraction of material 2. In this work, we take $\phi = 0$ without loss of generality. This choice leads to equal volume fractions for the two materials. $\mathcal{U}[a, b]$ denotes a uniform random distribution between a and b . $D \in [0, 1]$ is

the disorder parameter that determines the overall variation in grain size, with $D = 0$ leading to a periodic microstructure with no randomness. Within any grain, any wave q of the following form is a solution to (1):

$$v(x, t) = q(x \pm ct), \quad c = \sqrt{\frac{E}{\rho}} \quad (4)$$

where c is the wave speed and $v = \partial u / \partial t$ is the velocity of a material point. In the expression $x \pm ct$, the $+$ sign is for a “left-running” wave and the $-$ is for a “right-running” wave. Because the system is linear, arbitrary superpositions of waves in both directions are also solutions to (1). In the direct numerical simulation (DNS) technique that is used in this study, the waves that are superposed are composed of finite jumps in velocity, so that $q = H$ in (4), where H denotes the Heaviside step function.

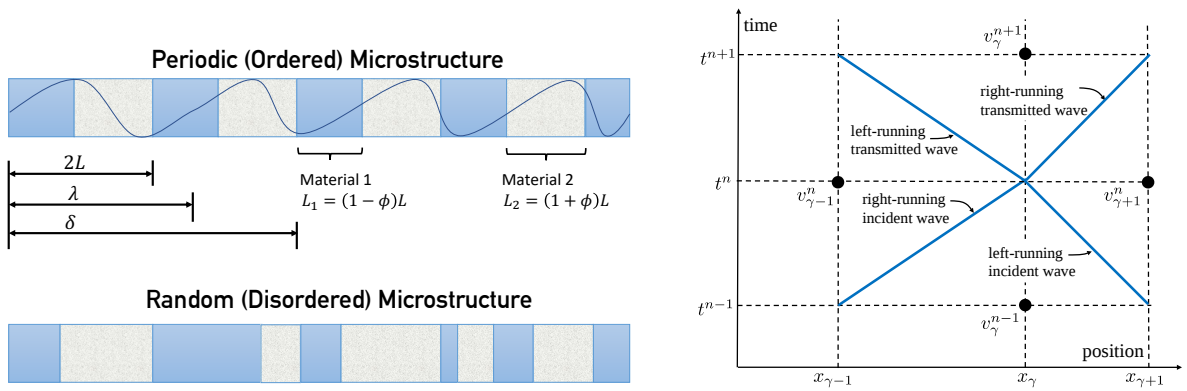


Fig. 2. Left: One-dimensional heterogeneous bars composed of materials 1 and 2. The oscillatory curve represents a moving wave with wavelength λ . The horizon δ for the nonlocal continuum model is shown. Left Top: Periodic microstructure with period $2L$. Left Bottom: Random microstructure. Right: Interaction of two waves in the DNS method.

Later in the paper, the group velocity $v_g(\omega)$ will be used, where ω is the angular frequency of a wave. In a homogeneous medium, the group velocity is defined by $v_g(\omega) = \frac{d\omega}{dk}$, where $k = 2\pi/\lambda$ is the wave number. With the DNS solver or any other simulation method, the group velocity can be estimated from the velocity of a wave packet with nominal frequency ω (see Figure 3). In both periodic and random bars, the qualitative dependence of v_g on ω is similar. Both have the same large-wavelength (low frequency) dependence, because $\lambda \gg L$ for long waves. In particular, both v_g functions have the same curvature as $\omega \rightarrow 0$. At smaller wavelengths, there are differences in

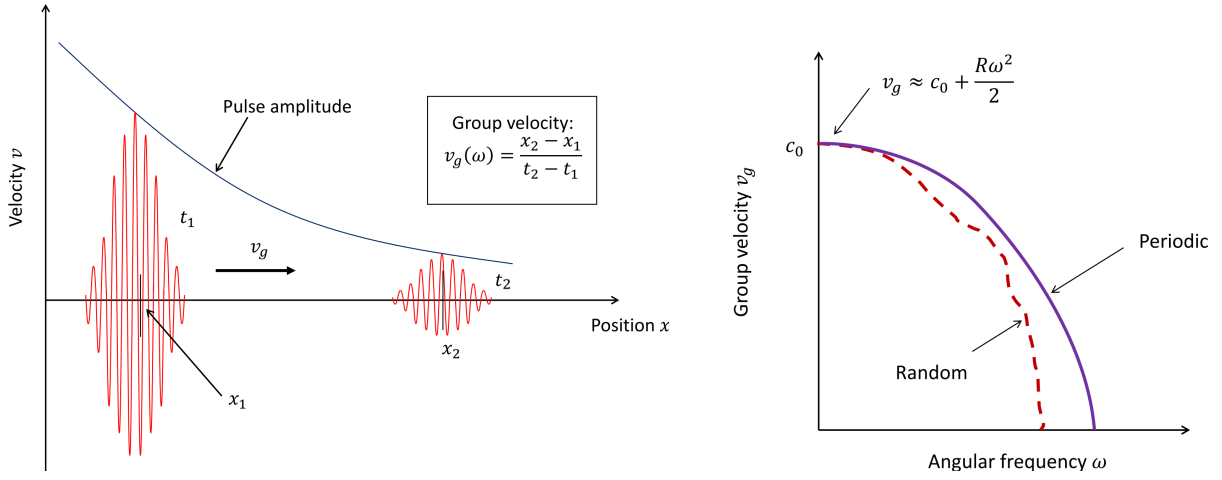


Fig. 3. Left: Group velocity estimation from the speed of a wave packet, which can attenuate significantly in random media. Right: Qualitative dependence of group velocity on frequency.

v_g between periodic and random systems. The method for evaluating v_g from wave packets is also susceptible to error near the band stop, because in this limit, the wave packet attenuates to nearly undetectable amplitudes as it propagates.

The goal of our learning algorithm is to learn a surrogate model that is able to predict wave propagation over distances that are much larger than the size of the microstructure and provide error estimation. For both training and validation purposes, we first generate high-fidelity data by solving the classical wave equation within a detailed model of the microstructure using the DNS solver, which will now be briefly described. For a given forcing term $f(x, t)$ and boundary and initial conditions, the DNS solver provides a solution for the velocity field $v(x, t)$. The bar is discretized into nodes, $\{x_\gamma\}$, such that it takes a constant amount of time Δt_{DNS} for a wave to travel through the cell between x_γ and $x_{\gamma+1}$, regardless of the elastic wave speed in the material between these two nodes (see the right plot of Figure 2 and (You et al. 2021a) for further details). The interval between two adjacent nodes is called a “cell.” Each cell is composed entirely of one material or the other, with elastic modulus E_1 or E_2 . These materials have wave speeds c_1 and c_2 , given by (4).

The right plot of Figure 2 shows left-running and right-running waves on an x - t diagram (also called a wave diagram). Because the wave speed varies between cells, the slopes shown in the

diagram also vary. Because the wave transit time within any cell must be constant, the spacing between DNS nodes varies along the bar. The DNS solver works by assuming that in each time step, a step wave travels from any node γ to its neighbor on the left or right. These neighbors also send waves to node γ . Within any cell, the two waves obey the jump condition

$$[\sigma] = \pm \rho c [v], \quad (5)$$

which is a consequence of the momentum balance. Note that the jump condition (5) applies to the moving waves, while (2) applies at fixed points. From these conditions, the DNS solver computes the updated material velocity v_γ^{n+1} explicitly from the values at the adjacent nodes in time step n . Details of the DNS solver can be found in (Silling 2021; You et al. 2021a).

After the velocity v_γ^{n+1} is computed, the displacement is updated by integrating the velocity over time:

$$u_{DNS}(x_\gamma, t^{n+1}) \approx u_\gamma^{n+1} = u_\gamma^n + \Delta t_{DNS} v_\gamma^{n+1}.$$

The displacements are mapped onto a mesh with constant spacing by interpolation between the irregularly spaced DNS node positions. The mapped displacements are used as training and validation data as described below.

We note that the DNS solver generates an exact wave velocity, allowing us to model the propagation of waves through many thousands of microstructural interfaces. The associated displacement is computed approximately and then interpolated onto the high-fidelity grid; this process potentially introduces numerical errors. Despite this fact, we consider these errors negligible and treat the interpolated displacements as the ground-truth measurements.

BACKGROUND AND RELATED MATHEMATICAL FORMULATION

In this section we introduce the nonlocal surrogate model used in this work and the proposed Bayesian learning approach. Consider the simulation on a spatial domain Ω and time domain $[0, T]$, and S observations of forcing terms $f^s(x, t)$ and their corresponding high-fidelity solution and/or experimental measurements of displacement fields $u_{DNS}^s(x, t)$. In the following we employ

these solutions as the ground-truth dataset, denoted as $\mathcal{D} := \{u_{DNS}^s\}_{s=1}^S$. Here we assume that both f^s and u_{DNS}^s measurements are provided on a set of time instances $t^n \in [0, T]$ and discretization points $x_i \in \Omega$. Without loss of generality, we assume that all measurements are provided on uniformly spaced spatial and time instances, with fixed spatial grid size Δx and time step size Δt , and we denote the collection of all discretization points as $\chi = \{x_i\}_{i=1}^L$. The overall goal of nonlocal operator regression is to provide a nonlocal model surrogate for the simulation of wave propagation in heterogeneous materials. Thus, we claim that given the same forcing terms $f^s(x, t)$, the corresponding solution, $u_{NL}^s(x, t)$, $(x, t) \in \Omega \times [0, T]$, of the nonlocal surrogate model provides a good approximation of the ground-truth data, i.e., $u_{NL}^s(x, t) \approx u_{DNS}^s(x, t)$.

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

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

which can be interpreted as a numerical approximation of the $L^2(\Omega \times [0, T])$ norm of u . Finally, in what follows, $\mathbf{I}_p$ denotes the $p \times p$ identity matrix; with an abuse of notation, when there is no confusion on its dimension, we simply use $\mathbf{I}$ to denote the identity matrix.

Nonlocal Operator Regression: learning nonlocal kernels

In this section we review the general nonlocal operator regression (NOR) approach developed in (You et al. 2021a). The goal of NOR is to find a nonlocal model that best describes the evolution of a homogenized quantity such as the propagation of stress waves in highly heterogeneous materials.

NOR starts from the assumption that a high-fidelity data set, satisfying an underlying high-fidelity model, is available. Given a force loading term $f(x, t)$, $(x, t) \in \Omega \times [0, T]$, proper boundary conditions and initial conditions, we represent the high-fidelity (HF) model as:

$$\frac{\partial^2 u_{HF}}{\partial t^2}(x, t) - \mathcal{L}_{HF}[u_{HF}](x, t) = f(x, t). \quad (6)$$

Here, $\mathcal{L}_{HF}$ is the HF operator that accounts for the detailed microstructure, and $u_{HF}(x, t)$ is the HF solution which can be provided either from fine-scale simulations or from experimental measurements in practice. In this work, we take u_{HF} as the numerical solution generated from the DNS solver as described in Section “Dispersion in Heterogeneous Materials”, i.e., $u_{HF} = u_{DNS}$. Analogously, we will refer to the homogenized effective nonlocal (NL) model as the homogenized surrogate, and assume it has the form

$$\frac{\partial^2 u_{NL}}{\partial t^2}(x, t) - \mathcal{L}_{NL}[u_{NL}](x, t) = f(x, t), \quad (7)$$

where the operator $\mathcal{L}_{NL} := \mathcal{L}_{K_C}$ is an integral operator associated with a nonlocal kernel K_C :

$$\mathcal{L}_{K_C}[u](x, t) = \int_{\Omega} K_C(x, y)(u(y, t) - u(x, t))dy, \quad (8)$$

and $u_{NL}(x, t)$ is the nonlocal solution. Here $\mathbf{C} := \{C_m\}_{m=0}^M \in \mathbb{R}^{M+1}$ is the parameter set of kernel K_C , which will be optimized during training. As shown in (Du et al. 2017), the second-order-in-time nonlocal equation in (7) is guaranteed to be well-posed as far as the kernel K_C is uniformly Lipschitz continuous. That means, the resultant surrogate model is guaranteed to be solvable in applications, when proper boundary conditions and numerical discretization methods are employed. Following (You et al. 2021b; You et al. 2022; You et al. 2021a), we choose to take K_C as a radial, sign-changing, nonlocal kernel function, compactly supported on the ball of radius δ centered at x , i.e., $B_{\delta}(x)$. Then, we parameterize the nonlocal kernel K_C as a linear combination of Bernstein basis polynomials:

$$K_C\left(\frac{|\xi|}{\delta}\right) = \sum_{m=0}^M \frac{C_m}{\delta^{d+2}} B_{m,M}\left(\left|\frac{\xi}{\delta}\right|\right), \quad (9)$$

where $d = 1$ is the dimension of the physical domain, Ω , and the Bernstein basis functions are defined as

$$B_{m,M}(\xi) = \binom{M}{m} \xi^m (1 - \xi)^{M-m} \quad \text{for } 0 \leq \xi \leq 1.$$

This choice ensures that the learnt model can be readily applied in simulation problems.

To obtain the optimal nonlocal surrogate operator $\mathcal{L}_{NL}$, the nonlocal operator regression approach seeks the best $\mathbf{C}$ through an optimization-based procedure. For each forcing term $f^s(x, t)$, let $u_{NL, \mathbf{C}}^s(x, t)$, $(x, t) \in \Omega \times [0, T]$ be the nonlocal solution of (7) corresponding to a specific set of kernel parameter $\mathbf{C} := \{C_m\}_{m=0}^M$. We aim to find the optimal parameter set $\mathbf{C}$ such that the (approximated) nonlocal solution $u_{NL, \mathbf{C}}^s(x_i, t^n)$ for a common loading f^s is as close as possible to the HF solution $u_{DNS}^s(x_i, t^n)$, for all provided observation pairs $\{(u_{DNS}^s, f^s)\}_{s=1}^S$ and $x_i \in \chi$, $t^n = 0, \dots, T/\Delta t$. To do so, for each given forcing term $f^s(x, t)$ we need to provide a numerical approximation of its corresponding nonlocal solution, $(u_{NL, \mathbf{C}}^s)_i^n \approx u_{NL, \mathbf{C}}^s(x_i, t^n)$. Here, with the initial condition $(u_{NL, \mathbf{C}}^s)_i^0 := u_{DNS}^s(x_i, 0)$ and proper boundary conditions, we discretize (7) with the central difference scheme in time and Riemann sum approximation of the nonlocal operator in space, and obtain the approximated nonlocal solution at time t^{n+1} as follows:

$$\begin{aligned} (u_{NL, \mathbf{C}}^s)_i^{n+1} &:= 2(u_{NL, \mathbf{C}}^s)_i^n - (u_{NL, \mathbf{C}}^s)_i^{n-1} + \Delta t^2 f^s(x_i, t^n) + \Delta t^2 \left(\mathcal{L}_{K_C, h}[u_{NL, \mathbf{C}}^s] \right)_i^n \\ &= 2(u_{NL, \mathbf{C}}^s)_i^n - (u_{NL, \mathbf{C}}^s)_i^{n-1} + \Delta t^2 f^s(x_i, t^n) \\ &\quad + \Delta t^2 \Delta x \sum_{x_j \in B_\delta(x_i) \cap \chi} K_{\mathbf{C}}(|x_j - x_i|) ((u_{NL, \mathbf{C}}^s)_j^n - (u_{NL, \mathbf{C}}^s)_i^n), \end{aligned} \quad (10)$$

where $\mathcal{L}_{K_C, h}$ is an approximation of $\mathcal{L}_{K_C}$ by the Riemann sum with uniform grid spacing Δx . The optimal parameters $\mathbf{C}^* = \{C_m^*\}$ and the corresponding nonlocal surrogate operator $\mathcal{L}_{NL} := \mathcal{L}_{K_{\mathbf{C}^*}}$ can be obtained by considering the normalized squared-loss of displacement in $\Omega \times [0, T]$:

$$\mathbf{C}^* = \underset{\mathbf{C}}{\operatorname{argmin}} \sum_{s=1}^S \frac{\|u_{NL, \mathbf{C}}^s - u_{DNS}^s\|_{L^2(\Omega \times [0, T])}^2}{\|u_{DNS}^s\|_{L^2(\Omega \times [0, T])}^2} + \lambda \|\mathbf{C}\|_{l^2}^2, \quad (11)$$

$$\text{s.t. } \mathcal{L}_{K_C} \text{ satisfies physics-based constraints.} \quad (12)$$

Here λ is a regularization parameter, and (12) depends on the partial physical knowledge of the heterogeneous material, which we will discuss later on in Section “Physics Constraints”.

Initialization for Parameters and Hyper-parameters

We note that the solution of (10) has numerical errors accumulated from $t = 0$ to $t = T$, and hence in practice the optimization problem (11) aims to minimize the accumulated error of displacement fields. This setting was originally proposed and employed in (You et al. 2021a), where the authors found that minimizing the accumulated error would help to learn physically-stable surrogate models, and such a stability plays a critical role in long-term prediction tasks. On the other hand, we also point out that in some application scenarios one can also choose to minimize the step-by-step time integration error. In (Lu et al. 2022; Zhang et al. 2022), an approximated nonlocal solution with one-step temporal error was considered:

$$\begin{aligned} (\tilde{u}_{NL,C}^s)_i^{n+1} &:= 2(u_{DNS}^s)_i^n - (u_{DNS}^s)_i^{n-1} + \Delta t^2 f^s(x_i, t^n) + \Delta t^2 (\mathcal{L}_{K_C,h}[u_{DNS}^s])_i^n \\ &= 2(u_{DNS}^s)_i^n - (u_{DNS}^s)_i^{n-1} + \Delta t^2 f^s(x_i, t^n) + \Delta t^2 \Delta x \sum_{x_j \in B_\delta(x_i) \cap \mathcal{X}} K_C(|x_j - x_i|) ((u_{DNS}^s)_j^n - (u_{DNS}^s)_i^n), \end{aligned}$$

where $(u_{DNS}^s)_i^n := u_{DNS}^s(x_i, t^n)$ represents the high-fidelity solution. A step-by-step loss of displacement is then formulated as

$$\tilde{\mathbf{C}}^* = \underset{\mathbf{C}}{\operatorname{argmin}} \sum_{s=1}^S \frac{\|\tilde{u}_{NL,C}^s - u_{DNS}^s\|_{l^2(\Omega \times [0,T])}^2}{\|u_{DNS}^s\|_{l^2(\Omega \times [0,T])}^2} + \lambda \|\mathbf{C}\|_{l^2}^2, \quad (13)$$

$$\text{s.t. } \mathcal{L}_{K_C} \text{ satisfies physics-based constraints.} \quad (14)$$

The loss function in (13) is in fact a quadratic equation with respect to $\mathbf{C}$. As such, NOR is equivalent to a linear regression model and solving (13) becomes a trivially linear problem. This fact makes hyper-parameter tuning, such as the selection of the regularization parameter λ , more efficient in (13). For instance, one can identify the optimal λ as the maximizer of the curvature of the curve, following the L-curve method (Hansen 2000; Lang and Lu 2022; Lu et al. 2022). Let l be a parametrized curve in $\mathbb{R}^2$, satisfying

$$l(\lambda) = (\alpha(\lambda), \beta(\lambda)) := (\log(\mathcal{E}(\mathbf{C}^*(\lambda))), \log(\mathcal{R}(\mathbf{C}^*(\lambda)))), \quad (15)$$

where $\mathcal{E}(\mathbf{C}^*(\lambda)) := \sum_{s=1}^S \frac{\|\tilde{u}_{NL, \mathbf{C}^*(\lambda)}^s - u_{DNS}^s\|_{L^2(\Omega \times [0, T])}^2}{\|u_{DNS}^s\|_{L^2(\Omega \times [0, T])}^2}$ is the loss function without regularization, and $\mathcal{R}(\mathbf{C}^*(\lambda))$ is the regularization term, which is taken as $\|\mathbf{C}^*(\lambda)\|_2^2$ in (13). Here, we used $\mathbf{C}^*(\lambda)$ to highlight the fact that the optimal parameter, $\mathbf{C}^*$, depends on the choice of the regularization parameter λ . The optimal parameter for λ then taken as the maximizer of the curvature of l as

$$\lambda^* = \operatorname{argmax}_{\lambda} \frac{\alpha' \beta'' - \alpha' \beta''}{(\alpha'^2 + \beta'^2)^{3/2}}. \quad (16)$$

This optimal parameter λ^* balances the loss $\mathcal{E}$ and the regularization. For further details of the L-curve method and discussions, we refer interested readers to (Hansen 2000).

In this work, we use the step-by-step loss formulation in (13) to provide an efficient estimation for $\mathbf{C}$ and to select the optimal regularization parameter λ . Then, the selected parameter set and hyperparameter will be employed in the optimization problem with accumulated loss, as an initial guess for the optimization solver and an estimated regularization parameter, respectively. Further details and discussions will be provided in Section “A Two-Phase Learning Algorithm”.

Physics Constraints

As illustrated in (You et al. 2021a), when some physical knowledge is available, such as the effective wave speed for infinitely long wavelengths and the curvature of the dispersion curve in the low-frequency limit, this knowledge can be incorporated into the optimization problem as physics-based constraints in (12) and (14). In particular, when the effective wave speed for infinitely long wavelengths, c_0 , is available, the corresponding physics-based constraint is:

$$\int_0^\delta \xi^2 K_{\mathbf{C}}(|\xi|) d\xi = \rho c_0^2, \quad (17)$$

where ρ is the effective material density. Discretizing (17) by Riemann sum, we obtain the first constraint on $\{C_m\}$:

$$\rho c_0^2 = \sum_{m=0}^M C_m \sum_{\eta=1}^{\lfloor \delta/\Delta x \rfloor} \frac{\eta^2 \Delta x^3}{\delta^3} B_{m,M} \left(\left| \frac{\eta \Delta x}{\delta} \right| \right) = \sum_{m=0}^M C_m A_{1m} \quad (18)$$

where $A_{1m} := \sum_{\eta=1}^{\lfloor \delta/\Delta x \rfloor} \frac{\eta^2 \Delta x^3}{\delta^3} B_{m,M} \left(\left| \frac{\eta \Delta x}{\delta} \right| \right)$. Furthermore, when the curvature of the dispersion curve in the low-frequency limit, R , is also available, the corresponding physics-based constraint is:

$$\int_0^\delta \xi^4 K_C(|\xi|) d\xi = -4\rho c_0^3 R. \quad (19)$$

Discretizing (19) with the Riemann sum approximation yields the second constraint on $\{C_m\}$:

$$-4\rho c_0^3 R = \sum_{m=0}^M C_m \sum_{\eta=1}^{\lfloor \delta/\Delta x \rfloor} \frac{\eta^4 \Delta x^5}{\delta^3} B_{m,M} \left(\left| \frac{\eta \Delta x}{\delta} \right| \right) = \sum_{m=0}^M C_m A_{2m} \quad (20)$$

where $A_{2m} := \sum_{\eta=1}^{\lfloor \delta/\Delta x \rfloor} \frac{\eta^4 \Delta x^5}{\delta^3} B_{m,M} \left(\left| \frac{\eta \Delta x}{\delta} \right| \right)$. These two physics-based equations are imposed as linear constraints on $\{C_m\}$. In this work, we consider the heterogeneous bar composed of alternating layers of two dissimilar materials, with (averaged) layer sizes $L_1 = (1 - \phi)L$, $L_2 = (1 + \phi)L$ for materials 1 and 2, respectively. Then the effective material density, Young's modulus and the wave speed are given by $\rho = ((1 - \phi)\rho_1 + (1 + \phi)\rho_2)/2$, $E = 2/((1 - \phi)E_1^{-1} + (1 + \phi)E_2^{-1})$, and $c_0 = \sqrt{E/\rho}$. R is the second derivative of the wave group velocity with respect to frequency at $\omega=0$. As discussed in Section “Dispersion in Heterogeneous Materials”, the group velocity is estimated from the speed of wave packets using the DNS solver. For the periodic microstructure, the curvature R of the function $v_g(\omega)$ at $\omega = 0$ is found by numerically differentiating this function. The value obtained for the present bar composition is $R = \frac{d^2 v_g^{DNS}}{d\omega^2}(0) = -0.006135$. The same value of R is also employed for the random microstructure because, as discussed in Section “Dispersion in Heterogeneous Materials”, the same dispersion properties apply to both periodic and random media with the same ϕ and L if the wavelength is much greater than the microstructural length scale.

When applying (18) and (20) in the optimization problem (11) or (13), we reformulate this constrained optimization problem such that an unconstrained optimization problem is obtained. In particular, denoting $\mathbf{C}_L := [C_0, \dots, C_{M-2}]^T$ and $\mathbf{C}_R := [C_{M-1}, C_M]^T$, (18) and (20) can be

rewritten as

$$\mathbf{A}_L \mathbf{C}_L + \mathbf{A}_R \mathbf{C}_R = \begin{bmatrix} \rho c_0^2 \\ -4\rho c_0^3 R \end{bmatrix}, \text{ where } \mathbf{A}_L := \begin{bmatrix} A_{10} & \cdots & A_{1,M-2} \\ A_{20} & \cdots & A_{2,M-2} \end{bmatrix}, \mathbf{A}_R := \begin{bmatrix} A_{1,M-1} & A_{1,M} \\ A_{2,M-1} & A_{2,M} \end{bmatrix}.$$

Hence, one can eliminate $\mathbf{C}_R$ by writing it as a linear expression in $\mathbf{C}_L$:

$$\mathbf{C}_R = \mathbf{A}_R^{-1} \left(\begin{bmatrix} \rho c_0^2 \\ -4\rho c_0^3 R \end{bmatrix} - \mathbf{A}_L \mathbf{C}_L \right), \quad (21)$$

and substituting into the optimization problem (11) or (13). Therefore, in the later sections we only need to solve for $\mathbf{C}_L$, and will demonstrate the algorithm for the unconstrained problem.

BAYESIAN NONLOCAL OPERATOR REGRESSION

Given observations of forcing terms $f^s(x, t)$, the corresponding DNS solution of displacement fields at time instance $t^n \in [0, T]$, and discretization points $x_i \in \chi$, in this section we formulate the Bayesian inference problem. [Here, we stress that the proposed learning approach does not require the knowledge of the high-fidelity model, but only the availability of high-fidelity data at sparse points in the computational domain. Thus, it is readily usable when only sparse experimental measurements are available.](#) Moreover, due to the fact that the target surrogate homogenized model is supposed to act at a larger scale without resolving the detailed microstructure, the model discrepancy would be the largest contributor to the overall modeling error and correspondingly the predictive uncertainty. For the s -th observation, we model the discrepancy between $u_{NL,C}^s(x_i, t^n)$ and the ground truth measurement $u_{DNS}^s(x_i, t^n)$ as additive independent unbiased Gaussian random noise ϵ , with:

$$u_{DNS}^s(x_i, t^n) = u_{NL,C}^s(x_i, t^n) + \epsilon_{s,i,n}, \quad \epsilon_{s,i,n} \sim \mathcal{N}(0, \tilde{\sigma}_s^2). \quad (22)$$

Here, with σ a constant independent of s , $\tilde{\sigma}_s := \sigma \cdot \left\| u_{NL,C}^s(x, t) \right\|_{l^2(\Omega \times [0, T])}$, and $\mathcal{N}(0, \tilde{\sigma}_s^2)$ represents the normal distribution with zero mean and standard deviation $\tilde{\sigma}_s$. Thus, σ is the standard deviation of $\left\{ u_{DNS}^s - u_{NL,C}^s \right\}_{s=1}^S$ after normalization w.r.t. the l^2 -norm of $\left\{ u_{NL,C}^s \right\}_{s=1}^S$. [Here, we point out](#)

that one of the ultimate goals is to use the learnt model as a homogenized surrogate to predict the material response under unseen loading and boundary/initial conditions and when high-fidelity measurements or models are not available. Thus, we choose to normalize with respect to the (computable) nonlocal surrogate solution instead of the (unlikely available or expensive to compute) high-fidelity data. For each observation, the likelihood function at (x_i, t^n) is

$$f(u_{DNS}^s(x_i, t^n) | \mathbf{C}) = \frac{1}{\sqrt{2\pi\tilde{\sigma}_s^2}} \exp\left(-\frac{|u_{NL,\mathbf{C}}^s(x_i, t^n) - u_{DNS}^s(x_i, t^n)|^2}{2\tilde{\sigma}_s^2}\right).$$

Therefore, given the independent noise construction, the likelihood for all the observations is

$$f(\mathcal{D} | \mathbf{C}) = \prod_{s,i,n=1}^{S,L,T/\Delta t} \left\{ \frac{1}{\sqrt{2\pi\sigma^2 \|u_{NL,\mathbf{C}}^s\|_{l^2(\Omega \times [0,T])}^2}} \exp\left(-\frac{|u_{NL,\mathbf{C}}^s(x_i, t^n) - u_{DNS}^s(x_i, t^n)|^2}{2\sigma^2 \|u_{NL,\mathbf{C}}^s\|_{l^2(\Omega \times [0,T])}^2}\right) \right\}. \quad (23)$$

We employ priors as a means of regularization. We use a multivariate normal prior distribution on the kernel parameter, $\mathbf{C} \sim \mathcal{N}(\hat{\mathbf{C}}, \frac{\hat{\sigma}^2}{\lambda} \mathbf{I})$. Here $\hat{\mathbf{C}}$ is the learnt parameter from the deterministic non-local operator regression technique introduced in Section “Background and Related Mathematical Formulation”, $\hat{\sigma}$ is defined as the σ_2 estimated in Step 1b of Algorithm 1 below, and λ is the chosen regularization parameter from the L-curve method, to achieve a good balance between the prior and the likelihood contributions. Since the normalized standard deviation of model discrepancy, σ , is anticipated to be smaller than 1, in our implementation we infer $\log(\sigma)$ instead of σ , and limit its range to $[-10, 0]$. Moreover, since we have no other prior knowledge about $\log(\sigma)$, we assume that its prior satisfies a uniform random distribution, i.e., $\log(\sigma) \sim \mathcal{U}[-10, 0]$. Then, combining the prior and likelihood in (23), we define the (unnormalized) posterior $\pi(\mathbf{C} | \mathcal{D}) \propto f(\mathcal{D} | \mathbf{C})P(\mathbf{C})$,

and obtain the associated negative log-posterior formulation [after eliminating the constant terms](#):

$$- \log(\pi(\mathbf{C}|\mathcal{D})) = \sum_{s=1}^S \left\{ \frac{\|u_{NL,\mathbf{C}}^s - u_{DNS}^s\|_{L^2(\Omega \times [0,T])}^2}{2\sigma^2 \|u_{NL,\mathbf{C}}^s\|_{L^2(\Omega \times [0,T])}^2} + N \log \left(\sigma \|u_{NL,\mathbf{C}}^s\|_{L^2(\Omega \times [0,T])} \right) \right\} + \lambda \frac{\|\mathbf{C} - \hat{\mathbf{C}}\|_{l^2}^2}{2\hat{\sigma}^2}, \quad (24)$$

where $N = L \frac{T}{\Delta t}$ is the total number of measurements for each observation pair, (f^s, u_{DNS}^s) .

A Two-Phase Learning Algorithm

To learn the posterior distribution (24), we employ an adaptive Markov chain Monte Carlo (MCMC) method ([Haario et al. 2001](#); [Andrieu et al. 2003](#); [Andrieu and Thoms 2008](#); [Debusschere et al. 2017](#)). MCMC is an effective applied Bayesian estimation procedure, generating random samples from the target posterior distribution. Particularly in high dimensional chains, MCMC benefits from a good initial guess to reduce burn-in. In order to accelerate convergence and obtain a good sampling of the posterior, we provide a good initial parameter estimate to start the chain.

To this end, a two-phase learning algorithm is proposed, with the main steps summarized in Algorithm 1. In this algorithm, an initialization phase, denoted as phase 1, is proposed before the MCMC algorithm in phase 2, with the purpose of providing good initial values that are close enough to the *maximum a posteriori* (MAP) parameter estimate from (24). [Through the initialization technique, one can also avoid manual parameter tuning.](#)

Phase 1 is composed of a sequence of deterministic optimization problems. First, in step **1a)** we consider the quadratic problem of minimizing the step-by-step error

$$\begin{aligned} \mathbf{C}_1 = \underset{\mathbf{C}}{\operatorname{argmin}} \quad & \sum_{s=1}^S \frac{\|\tilde{u}_{NL,\mathbf{C}}^s - u_{DNS}^s\|_{L^2(\Omega \times [0,T])}^2}{\|u_{DNS}^s\|_{L^2(\Omega \times [0,T])}^2} + \lambda \|\mathbf{C}\|_{l^2}^2, \\ \text{s.t. } \quad & \mathcal{L}_{K_C} \text{ satisfies physics-based constraints,} \end{aligned}$$

to provide estimates for the parameter solution, $\mathbf{C}_1$. Based on the estimated parameter, the approximated nonlocal solution $u_{NL,\mathbf{C}_1}^s$ can then be obtained following the numerical scheme (10), and an

Algorithm 1 A Two-Phase Learning Algorithm

1: Find a good initial state

1a) Learning $\mathbf{C}_1$ by minimizing the step-by-step error via:

$$\mathbf{C}_1 := \underset{\mathbf{C}}{\operatorname{argmin}} \sum_{s=1}^S \frac{\left\| \tilde{u}_{NL,\mathbf{C}}^s - u_{DNS}^s \right\|_{L^2(\Omega \times [0,T])}^2}{\left\| u_{DNS}^s \right\|_{L^2(\Omega \times [0,T])}^2} + \lambda \|\mathbf{C}\|_{l^2}^2. \quad (25)$$

Then calculate σ_1 as the standard deviation of $u_{NL,\mathbf{C}_1} - u_{DNS}$, and update the regularization parameter, λ_1 , following (29).

1b) With initial values $\mathbf{C}_1$ and regularization parameter λ_1 obtained from step **1a**), find $\mathbf{C}_2$ and σ_2 as follows:

$$(\mathbf{C}_2, \sigma_2) := \underset{\mathbf{C}, \sigma}{\operatorname{argmin}} \sum_{s=1}^S \left\{ \frac{\left\| u_{NL,\mathbf{C}}^s - u_{DNS}^s \right\|_{L^2(\Omega \times [0,T])}^2}{2\sigma^2 \left\| u_{DNS}^s \right\|_{L^2(\Omega \times [0,T])}^2} + N \log \left(\sigma \left\| u_{DNS}^s \right\|_{L^2(\Omega \times [0,T])} \right) \right\} + \lambda_1 \frac{\|\mathbf{C}\|_{l^2}^2}{2\sigma_1^2}. \quad (26)$$

1c) With the initial values $\hat{\mathbf{C}} := \mathbf{C}_2$, determine the regularization parameter by solving the linear regression problem with the quadratic loss function:

$$\mathbf{C}_3 := \underset{\mathbf{C}}{\operatorname{argmin}} \sum_{s=1}^S \frac{\left\| \tilde{u}_{NL,\mathbf{C}}^s - u_{DNS}^s \right\|_{L^2(\Omega \times [0,T])}^2}{\left\| u_{NL,\mathbf{C}}^s \right\|_{L^2(\Omega \times [0,T])}^2} + \lambda \|\mathbf{C} - \hat{\mathbf{C}}\|_{l^2}^2, \quad (27)$$

and update the regularization parameter λ_2 following (34).

2: Perform MCMC

With the initial values $(\mathbf{C}_2, \sigma_2)$ and the regularization parameter λ_2 from phase 1, run MCMC and sample the posterior $p(\mathbf{C}, \sigma | \mathcal{D})$, where the negative log-posterior, up to an additive constant, is given by

$$\sum_{s=1}^S \left\{ \frac{\left\| u_{NL,\mathbf{C}}^s - u_{DNS}^s \right\|_{L^2(\Omega \times [0,T])}^2}{2\sigma^2 \left\| u_{NL,\mathbf{C}}^s \right\|_{L^2(\Omega \times [0,T])}^2} + N \log \left(\sigma \left\| u_{NL}^s \right\|_{L^2(\Omega \times [0,T])} \right) \right\} + \lambda_2 \frac{\|\mathbf{C} - \mathbf{C}_2\|_{l^2}^2}{2\sigma_2^2}. \quad (28)$$

3: Postprocessing

Extract effective samples from the MCMC chain, and estimate statistical moments of the corresponding solutions and other quantities of interests.

estimate of the parameter σ , which is denoted as σ_1 , is provided by calculating the (normalized) standard deviation of $u_{NL,C_1} - u_{DNS}$. We also estimate the regularization parameter, λ_1 , as:

$$\lambda_1 := \sum_{s=1}^S \frac{\|u_{NL,C_1}^s - u_{DNS}^s\|_{L^2(\Omega \times [0,T])}^2}{\|u_{DNS}^s\|_{L^2(\Omega \times [0,T])}^2 \|C_1\|_{L^2}^2}. \quad (29)$$

Intuitively, this parameter guarantees that the accumulated error and the regularization term have similar scales. Here, we use the accumulated error instead of the step-by-step error, to be consistent with the MCMC formulation (24). Then, based on the estimated parameter C_1 and the regularization parameter λ_1 , in step **1b**) we solve a deterministic optimization problem:

$$(C_2, \sigma_2) := \underset{C, \sigma}{\operatorname{argmin}} \sum_{s=1}^S \left\{ \frac{\|u_{NL,C}^s - u_{DNS}^s\|_{L^2(\Omega \times [0,T])}^2}{2\sigma^2 \|u_{DNS}^s\|_{L^2(\Omega \times [0,T])}^2} + N \log \left(\sigma \|u_{DNS}^s\|_{L^2(\Omega \times [0,T])} \right) \right\} + \lambda_1 \frac{\|C\|_{L^2}^2}{2\sigma_1^2}, \quad (30)$$

$$\text{s.t. } \mathcal{L}_{K_C} \text{ satisfies physics-based constraints.} \quad (31)$$

The above formulation acts as an approximation of the closed form of the negative log-posterior formulation in (24). In fact, the first two terms in the loss function aim to provide an approximation of the first two terms in (24), except that the standard deviation is normalized with respect to the DNS data. The last term comes from the multivariate normal prior distribution assumption on the kernel parameter, which acts as a regularization. This step is solved with the L-BFGS method. Then, based on the estimated solution C_2 , in step **1c**) we set $\hat{C} := C_2$ consider the quadratic formulation which can be seen as an approximation of the first and last term in (24):

$$C_3 := \underset{C}{\operatorname{argmin}} \sum_{s=1}^S \frac{\|\tilde{u}_{NL,C}^s - u_{DNS}^s\|_{L^2(\Omega \times [0,T])}^2}{\|u_{DNS}^s\|_{L^2(\Omega \times [0,T])}^2} + \lambda \|C - \hat{C}\|_{L^2}^2, \quad (32)$$

$$\text{s.t. } \mathcal{L}_{K_C} \text{ satisfies physics-based constraints,} \quad (33)$$

and can be solved as a linear problem. Here, the L-curve method is again employed with the loss function defined as $\mathcal{E}(\mathbf{C}(\lambda)) := \sum_{s=1}^S \frac{\|\tilde{u}_{NL, \mathbf{C}(\lambda)}^s - u_{DNS}^s\|_{l^2(\Omega \times [0, T])}^2}{\|u_{DNS}^s\|_{l^2(\Omega \times [0, T])}^2}$ and the regularization term as $\mathcal{R}(\mathbf{C}(\lambda)) := \|\mathbf{C}(\lambda) - \hat{\mathbf{C}}\|_{l^2}^2$. With the estimated solution $\mathbf{C}_3$, we then update the regularization parameter for phase 2, which will be denoted as λ_2 :

$$\lambda_2 := \sum_{s=1}^S \frac{\|u_{NL, \mathbf{C}_3}^s - u_{DNS}^s\|_{l^2(\Omega \times [0, T])}^2}{\|u_{NL, \mathbf{C}_3}^s\|_{l^2(\Omega \times [0, T])}^2 \|\mathbf{C}_3 - \mathbf{C}_2\|_{l^2}^2}. \quad (34)$$

Finally, with the initial values $(\mathbf{C}_2, \sigma_2)$ and the regularization parameter λ_2 from phase 1, we apply adaptive MCMC using the log-posterior (24), where the pair $(\hat{\mathbf{C}}, \hat{\sigma})$ as well as the initial guess of $(\mathbf{C}, \sigma)$ are set as $(\mathbf{C}_2, \sigma_2)$. Our adaptive MCMC is implemented using the Uncertainty Quantification toolkit (UQTK) (Debusschere et al. 2004; Debusschere et al. 2017).

APPLICATION TO A HETEROGENEOUS ELASTIC BAR

In this section, we examine the efficacy of the proposed BNOR model and the two-phase learning algorithm, by considering the stress wave propagation problem described in Section “Dispersion in Heterogeneous Materials”. Here, we seek a nonlocal homogenized model for the stress wave propagation in one-dimensional heterogeneous bars. Two exemplar heterogeneous bars are considered, one with periodic microstructure and one with random microstructure. For this problem, the goal is to obtain an effective nonlocal surrogate model from ground truth datasets generated by the DNS solver, acting at a much larger scale than the size of the microstructure. Since this problem has no ground-truth nonlocal kernel, we evaluate the surrogate by measuring its effectiveness in reproducing DNS data in applications that are subject to different loading conditions with a much longer time than the problems used as training data. To directly examine the extent to which our surrogate model reproduces the dispersion properties in the heterogeneous material, we also compare the group velocity curves from our model with the curves computed with DNS. Finally, we require that the learnt surrogate should provide a physically stable material model. To check this, we report the dispersion curve, whose positivity indicates that the learnt nonlocal

model is physically stable. Besides the predicted MAP solutions, the Bayesian model also provides estimation of the predictive uncertainty for all above quantities of interests.

In this work, we use a NVIDIA 3060 GPU for all training tasks. In the initialization phase, phase 1a and phase 1c take a few seconds to complete and phase 1b takes approximately 1 hour. We run phase 2 for approximately 3 days to generate 300,000 MCMC samples, ensuring that enough (approximately 4,000) effective samples are obtained.

Example 1: A Bar with Periodic Microstructure

Data generation and settings. First, we consider a heterogeneous bar with a periodic microstructure. As illustrated in the left top plot in Figure 2, the bar is a layered medium composed of two components, with the size of each layer $L = 0.2$. Components 1 and 2 have the same density $\rho = 1$ and Young's moduli $E_1 = 1$ and $E_2 = 0.25$, respectively. For the purpose of training and validation, we generate the DNS dataset with three types of data, and use the first two for training and the last one for validation of our algorithm. For all data we set the discretization parameters for the DNS solver as $\Delta t_{DNS} = 0.01$, $\max\{\Delta x_{DNS}\} = 0.01$, and consider the symmetric domain $\Omega = [-b, b]$. In what follows, u represents the displacement, v the velocity, and f an external loading. The three types of data are chosen to follow a similar setting as in (You et al. 2021a):

Type 1: Oscillating source (20 observations). We set $b = 50$. The bar starts from rest such that $v(x, 0) = u(x, 0) = 0$, and an oscillating loading is applied with $f(x, t) = e^{-\left(\frac{2x}{5kL}\right)^2} e^{-\left(\frac{t-t_0}{t_p}\right)^2} \cos^2\left(\frac{2\pi x}{kL}\right)$ with $k = 1, 2, \dots, 20$. Here we take $t_0 = t_p = 0.8$.

Type 2: Plane wave with ramp (11 observations). We also set the domain parameter as $b = 50$. 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

$$v(-b, t) = \begin{cases} \sin(\omega t) \sin^2\left(\frac{\pi t}{30}\right), & t \leq 15 \\ \sin(\omega t), & t > 15 \end{cases}$$

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

Type 3: Wave packet (3 observations). We consider a longer bar with $b = 133.3$, 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 $v(-b, t) = \sin(\omega t) \exp(-(t/5 - 3)^2)$ with $\omega = 2, 3.9$, and 5 .

For all data types, 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 data type 3.

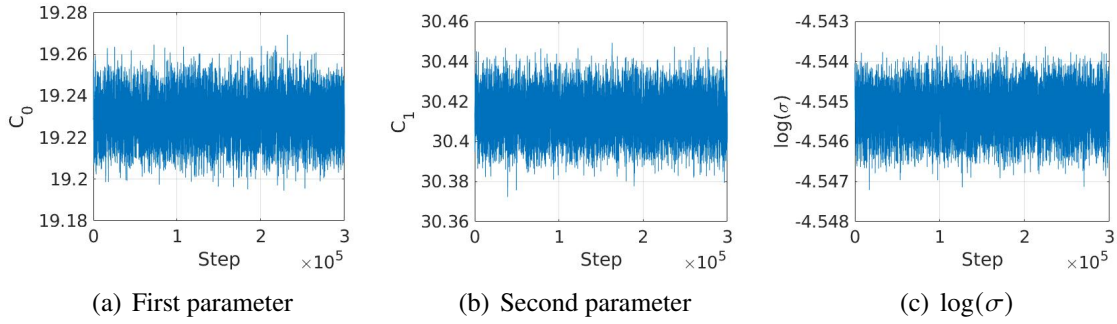


Fig. 4. The first two kernel parameters C_0 , C_1 and $\log(\sigma)$ for a bar with periodic microstructure, plotted against the iteration number after eliminating the burn-in stage (the trace plot of MCMC).

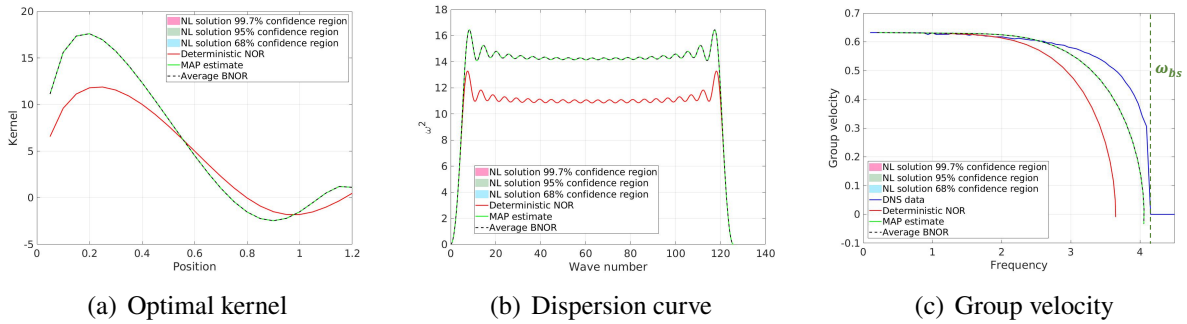


Fig. 5. Optimal kernel, group velocity and dispersion curve for a bar with periodic microstructure. Here, the confidence region almost coincides with the curve of average BNOR, since the uncertainty in the kernel parameters is low. We also point out that the results from MAP point estimate via minimizing (28) coincides with the averaged kernel from BNOR.

Results from Bayesian inference An MCMC chain with 300,000 steps is generated, with approximately 30% acceptance rate. To present the results, we post-process the chain based on 3,856

equally spaced samples, where the effective sample size (ESS) was calculated following the formulation in (Vats et al. 2019). To demonstrate the efficacy of our two-phase learning algorithm, we also perform an ablation study of running MCMC without the initialization estimated from phase 1. As shown in the Appendix Figure 14, the acceptance rate is very low (0.0059) and only a few proposals are sampled. Figure 4 shows a trace plot of the MCMC chain, illustrating good mixing of the chain. In Figure 5(a), the nonlocal kernels using the learnt parameters are demonstrated, with predictive uncertainty ranges of 68%, 95% and 99.7% provided. Here, the confidence ranges were calculated based on the push-forward of the marginal posterior on $\mathbf{C}$ (using the samples from MCMC), through the nonlocal model (7), and the averaged BNOR kernel is calculated by evaluating the posterior mean of the kernel parameter $\mathbf{C}$. Specifically, denoting the posterior mean of $\mathbf{C}$, as $\bar{\mathbf{C}}$, the averaged nonlocal kernel is evaluated as $K_{\bar{\mathbf{C}}} \left(\frac{|\xi|}{\delta} \right)$. Both the averaged nonlocal kernel from the proposed BNOR algorithm and the deterministic nonlocal kernel from the original NOR are reported for comparison, together with the MAP estimate. One can see that the averaged kernel from BNOR is close to the original NOR kernel when $\xi = |y - x|$ is relatively large (> 0.6) with an almost negligible predictive uncertainty, while the discrepancy becomes more significant when ξ approaches zero. However, the kernel values at $\xi \approx 0$ have a relatively small impact on the nonlocal operator $\mathcal{L}_{K_{\mathbf{C}}}$, and consequently also on the corresponding nonlocal solution u_{NL} as well as the loss function in our optimization problems. Moreover, sign-changing behaviors are observed on both kernels, which are consistent with literature (Xu and Foster 2020; You et al. 2021b). Then, in Figure 5(b) we illustrate the deterministic and mean estimates, and predictive uncertainty ranges, of the dispersion curve. Here, we can see that both the averaged model from BNOR and the deterministic model from NOR possess positivity, indicating that both models correspond to a family of physically stable material models. Finally, in Figure 5(c) we plot the estimated group velocity against frequency ω , in comparison with the group velocity profile from the DNS simulation. These results indicate that both the averaged model from BNOR and the deterministic model from original NOR are able to match the DNS behavior for low frequency values ω . When the frequency is getting closer to the band stop, ω_{bs} , comparing with the deterministic model from original NOR, we note

that the averaged model from BNOR has reproduced the group velocity of the DNS model with better accuracy. Here, we stress that since the method for evaluating the DNS v_g from wave packets is susceptible to error near the band stop, it is generally infeasible to expect our nonlocal surrogate model to match perfectly with the DNS group velocity. For both the dispersion curve and group velocity, predictive uncertainties are again negligible.

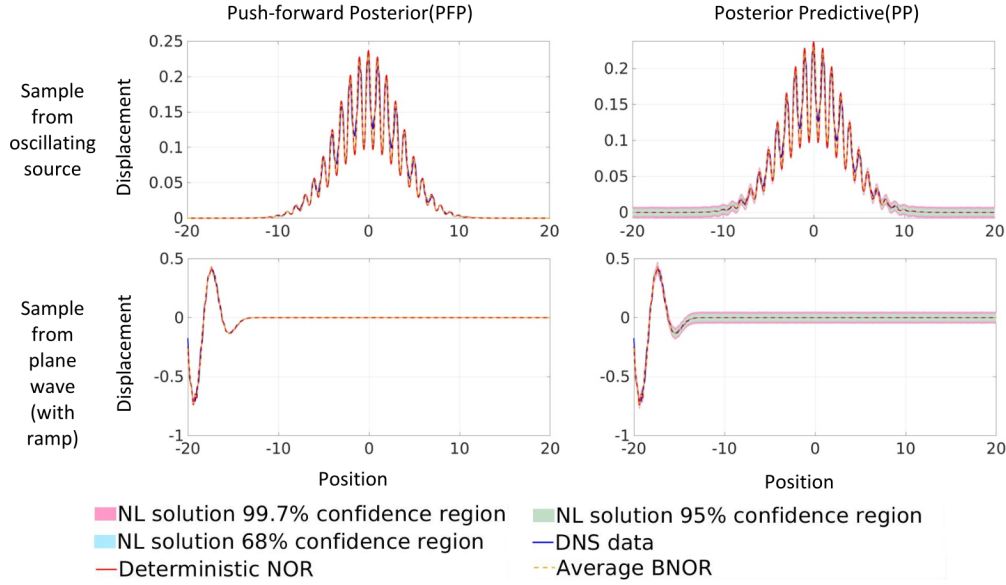


Fig. 6. Comparison of two posterior uncertainty results for the prediction of two exemplar examples from the training dataset, on the periodic microstructure case. Left: displacement field prediction and confidence regions from the push-forward posterior (PFP) approach. Right: displacement field prediction and confidence regions from the posterior predictive (PP) approach.

We note in particular that the evident negligible level of predictive uncertainty in these results is due to the fact that (1) our error model is rather simple, presuming independent noise, such that each additional data point adds information, and (2) we have a large number of data points. The combination of both factors results in the posterior on $\mathbf{C}$ being highly concentrated, thus exhibiting minimal uncertainty, even though the discrepancy between the two models is non-negligible. In order to endow the nonlocal predictive model with uncertainties that better approximate the discrepancy between it and the high-fidelity model, one can resort to a more elaborate statistical construction to represent model error (Kennedy and O’Hagan 2000), particularly as embedded in the model construction (Sargsyan et al. 2015; Sargsyan et al. 2019). We reserve this pathway

forward for future work.

We illustrate further, in Fig. 6, posterior uncertainty on two predicted solutions that are part of the training data. In each case, we highlight two posterior uncertainty results. The first, commonly termed the push-forward posterior (PFP), is the push-forward of the marginal posterior $p(\mathbf{C}|D)$ through the non-local model $u_{NL,\mathbf{C}}(x,t)$. The second, termed the posterior predictive (PP), is the push-forward of the full posterior $p(\mathbf{C}, \sigma|D)$ through the data model $u_{NL,\mathbf{C}}(x,t) + \epsilon$. The PFP provides the posterior uncertainty on the model predictions given posterior knowledge of $\mathbf{C}$. On the other hand, the PP provides a posterior density on predictions of the data model, which serves as a diagnostic measure of the quality of this model, composed of the physical model and the proposed error model, as a predictor of the data (Note that, in an additive model error framework (Kennedy and O'Hagan 2001) where the ϵ term, defined to include some correlation structure, would be included in "corrected" model, such that the associated PP & PFP are equivalent.). It is expected that the PP would exhibit higher uncertainty on predictions than the PFP. Further, an ideal PP should span the data, and its samples ought to be statistically indistinguishable from the data. Considering the results in Fig. 6, we can see that the BNOR solution with the mean $\mathbf{C}$, the deterministic NOR, and the DNS solution are all very well matched for the plane wave case, with discernible differences in the displacement magnitudes in the oscillating source case. Further, we see that the PFP results in negligible predictive uncertainty, as already observed in Fig. 5 for the kernel and associated quantities of interest. On the other hand, the PP results exhibit non-negligible uncertainty, particularly at low mean-output levels for positions beyond ± 10 for the oscillating source case, and above position -15 for the plane wave case. However, we find that the PP uncertainty is again negligible at high signal levels around position 0 in the oscillating source case, and below position -15 in the plane wave case. Moreover, the PP uncertainty is small relative to the discrepancy between the model and the data at large signal levels in the oscillating source case, while it is larger than the discrepancy at low signal levels. Clearly, the PP does not do an ideal job of spanning the discrepancy from the data, indicating that a more flexible/accurate, perhaps embedded, error model would be needed to better capture the data error.

Validation on wave packet In order to further test the predictive performance of our surrogate model, we now demonstrate its capability in reproducing DNS simulations through the prediction error of u on data type 3, involving wave packet problems where a long time simulation is conducted on a long bar, and the loading scenario is substantially different from the training cases. With the proposed BNOR approach, for each loading scenario both the prediction of displacement field $u_{NL,C}(x_i, t^n)$ and the 68% – 95% – 99.7% confidence regions are provided. Specifically, we present both the PFP $p(u_{NL,C}(x_i, t^n)|D)$ and PP $p(u_{NL,C}(x_i, t^n) + \epsilon|D)$ results, where $\epsilon \sim \mathcal{N}(0, \sigma^2 ||u_{NL,C}(x, t)||_{l_2([-b,b] \times [0,T])}^2)$.

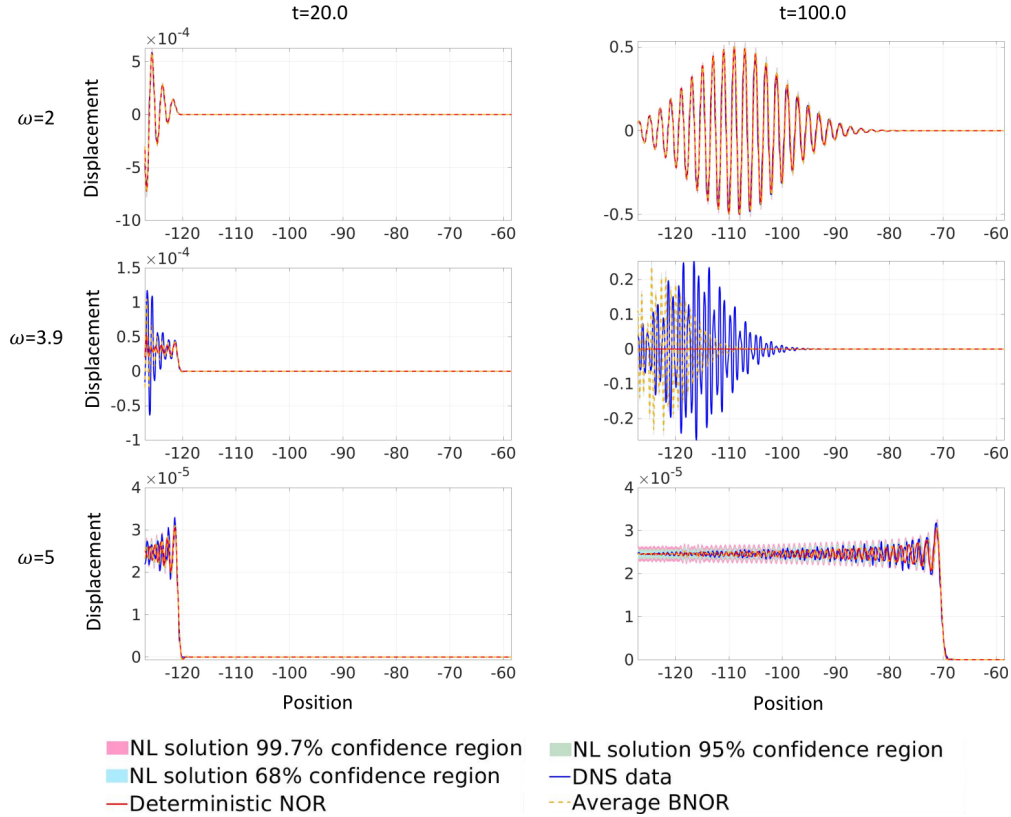


Fig. 7. Validation on the wave packet traveling problem for a bar with periodic microstructure, illustrating PFP uncertainty. Plots from top to bottom are corresponding to different loading frequencies: (top) $\omega = 2 < \omega_{bs}$, (middle) $\omega = 3.9 \approx \omega_{bs}$, and (bottom) $\omega = 5 > \omega_{bs}$. Left column shows the simulation results on a relatively short time ($T = 20$), and the right column demonstrates long time simulation results ($T = 100$).

In Figures 7 and 8 we consider solutions corresponding to three values of frequency ω : $\omega_1 = 2 < \omega_{bs}$, $\omega_2 = 3.9 \approx \omega_{bs}$ and $\omega_3 = 5 > \omega_{bs}$, respectively. Further, results in the two figures

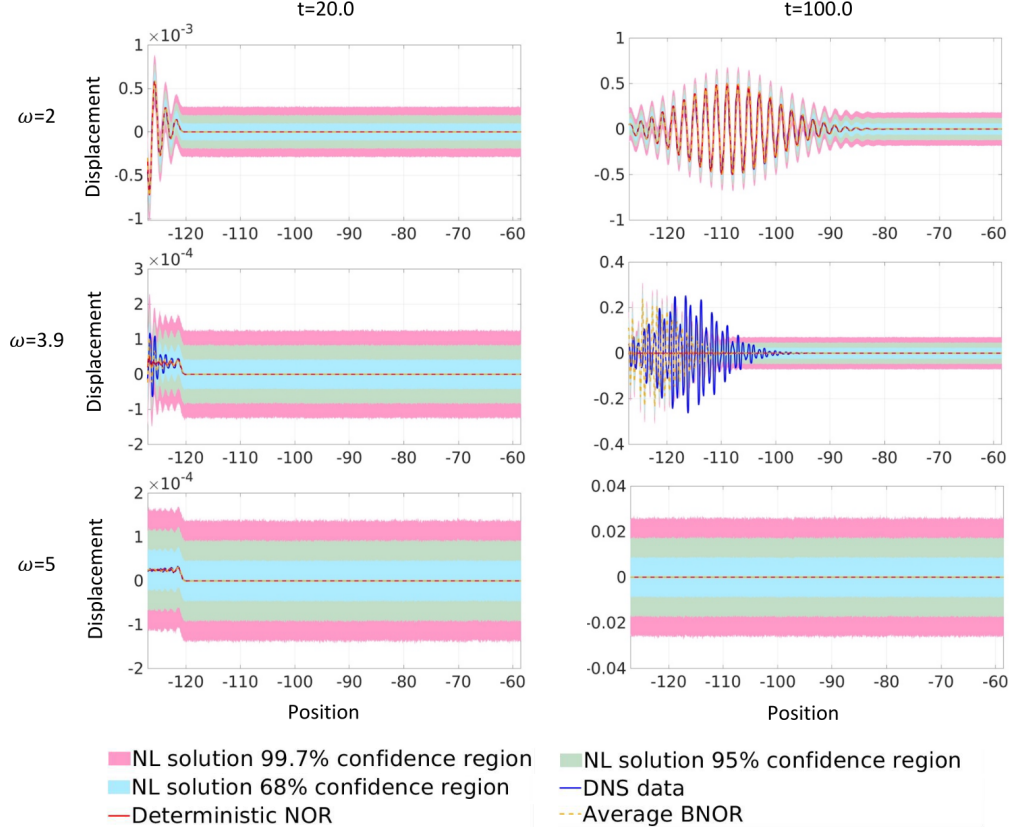


Fig. 8. Validation on the wave packet traveling problem for a bar with periodic microstructure, illustrating PP uncertainty. Plots from top to bottom are corresponding to different loading frequencies: (top) $\omega = 2 < \omega_{bs}$, (middle) $\omega = 3.9 \approx \omega_{bs}$, and (bottom) $\omega = 5 > \omega_{bs}$. Left column shows the simulation results on a relatively short time ($T = 20$), and the right column demonstrates long time simulation results ($T = 100$).

illustrate PFP (Fig. 7) and PP (Fig. 8) uncertainty as indicated. For the the first two values of ω , the exact stress wave is anticipated to be traveling in time. For the last case, since the value of ω is beyond the band stop and corresponds to a zero DNS group velocity, the exact wave does not travel in time. Both short time ($t = 20$) and long time ($t = 100$) predictions are considered for each case. The deterministic solution from the standard NOR algorithm and the DNS data are also reported for comparison. We observe at the outset the negligible PFP uncertainty (Fig. 7), as already observed in Fig. 6 for training data predictions, while larger PP uncertainty is evident (Fig. 8), and quite prominently as the displacement goes to zero.

Further, for the loading frequency $\omega = 2.0$ case, we observe good agreement between the mean of the nonlocal solutions and the DNS data for both $t = (20, 100)$ cases. Thus, the optimal

nonlocal surrogate corresponding to our learnt samples can accurately reproduce both short and long-time stress wave propagation in this frequency range. Here, we stress that although we refer to $t = 20$ as a relative short term, it is still much longer ($10\times$) than the total time interval our training algorithm has seen from the training samples. These findings illustrate the generalization property of our algorithm in extrapolation tasks in this frequency range. On the other hand, considering the wave propagation results for the loading frequency $\omega = 3.9$ case, which is close to the band stop frequency, a larger discrepancy between the DNS solution and the predicted displacement field is observed, especially in the region near the wave front. More specifically, at $t = 20$, the mean of the nonlocal solutions is slightly off from the DNS solution, while the latter still lies inside the 68% PP confidence region (Fig. 8). This discrepancy grows appreciably as time moves forward, as can be seen for the $t = 100$ case, extending well beyond the PP range. Clearly, our extrapolative behavior is poor for loading frequencies in the vicinity of the band stop frequency. Finally, when the loading frequency $\omega = 5.0 > \omega_{bs}$ and the wave barely propagates, as shown in both figures, our model successfully captures the phenomenon that the wave stops traveling.

Generally, based on these results, we can say that the proposed nonlocal surrogate model performs well in short term extrapolative prediction tasks, and the posterior predictive associated with the current additive error model provides an adequate coverage of the DNS data, even in extrapolation, when the loading frequency is away from the band stop frequency, ω_{bs} . To further improve the prediction accuracy and the model discrepancy representation, more training observations on a longer time and/or a more flexible and sophisticated error model employing a Gaussian process construction should help, allowing always for the lack of expected accuracy from the statistical model in extrapolation if the mean model itself is failing. We note in particular that the point-wise mismatch between discrepancy magnitude and predictive uncertainty from the calibrated model evident in the above figures is the expected result of the implicit mean-square averaging of the discrepancies/errors for all data points in the present additive *i.i.d.* Gaussian Likelihood construction. Our estimated standard deviation of the error, as well as our parametric uncertainties, being effectively the result of root-mean-squared averaging of these discrepancies, lead to predictive

uncertainty that is smaller than the largest discrepancies observed in the data, and larger than the smallest observed discrepancies.

Example 2: A Bar with Random Microstructure

Data generation and settings. We now consider a heterogeneous bar with random microstructure, where the size of each layer is now defined by a random variable. In particular, the density for both components is still set as $\rho = 1$ and the Young's moduli are $E_1 = 1$ and $E_2 = 0.25$. The layer sizes L_1, L_2 are two random variables, both satisfying a uniform distribution: $L_1, L_2 \sim \mathcal{U}[(1 - D)L, (1 + D)L]$, with the averaged layer size $L = 0.2$ and the disorder parameter $D = 0.5$. To generate the training dataset, we simulate the wave propagation using the DNS solver, under the same settings as in data type 1 and data type 2 of the periodic bar case. Then, the wave packet problem is again considered for the purpose of validation. Comparing with the periodic microstructure case, from the group velocity generated by the DNS simulations (see Figure 3 right plot) we note that the band stop generally occurs at a lower frequency in the random microstructure case. In fact, for the microstructure considered in this bar, an estimated band stop frequency $\omega_{bs} \approx 3$ can be obtained from the DNS simulations. Thus, in this section, wave packets with frequencies $\omega = 1, 2, 3$ and 4 are considered as the validation samples, 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} .

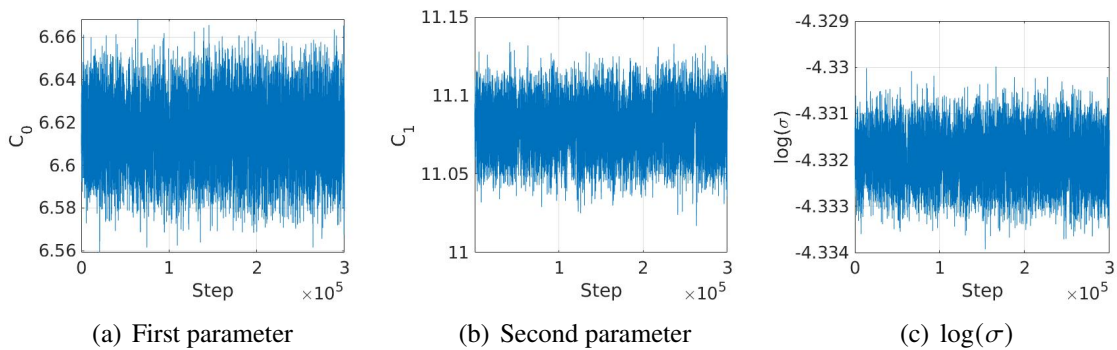


Fig. 9. The first two kernel parameters C_0, C_1 and $\log(\sigma)$ for a bar with random microstructure, plotted against the iteration number after eliminating the burn-in stage (the trace plot of MCMC).

Results from Bayesian inference. We run an MCMC chain with 300,000 steps, with a 31% acceptance rate, and post-process it using 4,212 equally spaced samples guided by an estimated ESS. The ESS, the trace plot, the predicted averaged kernel, and its dispersion properties are all calculated similarly as for the periodic bar. Figure 9 shows the exemplar trace plot of the MCMC chain for two kernel parameters, C_0 and C_1 , and σ , illustrating good mixing. Compared with the periodic bar, slightly larger values of σ are obtained here. Hence, the randomness in material microstructure is anticipated to induce a larger model discrepancy. Then, in Figure 10 we plot the

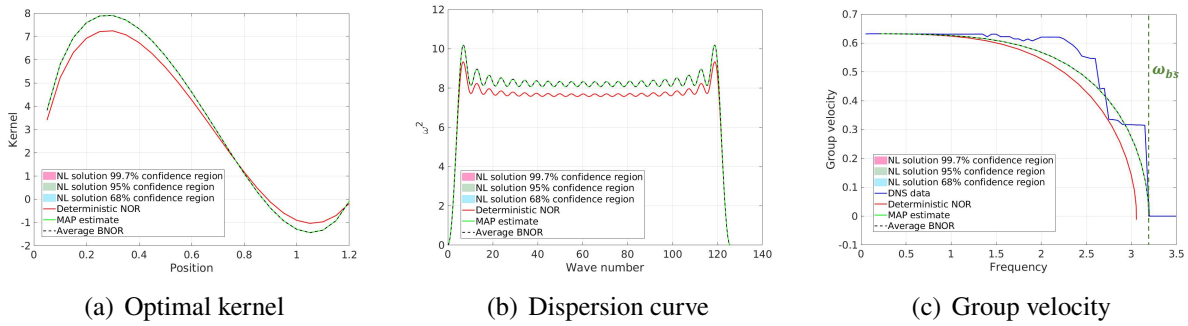


Fig. 10. Optimal kernel, group velocity and dispersion curve for a bar with random microstructure. Since the uncertainty in the kernel parameters remains low, the confidence regions generally coincide with the curve of average BNOR. *Similar as in the periodic bar case, here the results from MAP point estimate via minimizing (28) again coincides with the averaged kernel from BNOR.*

predicted averaged nonlocal kernel, the dispersion curve, and the group velocity profiles, together with the 68% – 95% – 99.7% PFP confidence regions. The results from original NOR algorithm and the MAP estimate are also reported for comparison. For the estimated nonlocal kernels, as demonstrated in Figure 10(a) we observe a small discrepancy between the averaged kernel and the deterministic kernel from NOR, together with very small confidence regions. These trends are also observed in Figure 10(b) and Figure 10(c), for the dispersion curves and group velocity profiles, respectively. As in the periodic bar case, for this random microstructure the predicted nonlocal model from BNOR again possesses physical stability, and successfully identifies the band stop. In Figure 10(c), a relatively larger discrepancy is observed between the group velocity from the DNS solver and profile from the estimated nonlocal surrogate, possibly due to the fact that the material

randomness introduces larger errors in the approximated DNS group velocity, v_g , especially for the frequencies near the band stop, as discussed in Section “Dispersion in Heterogeneous Materials”.

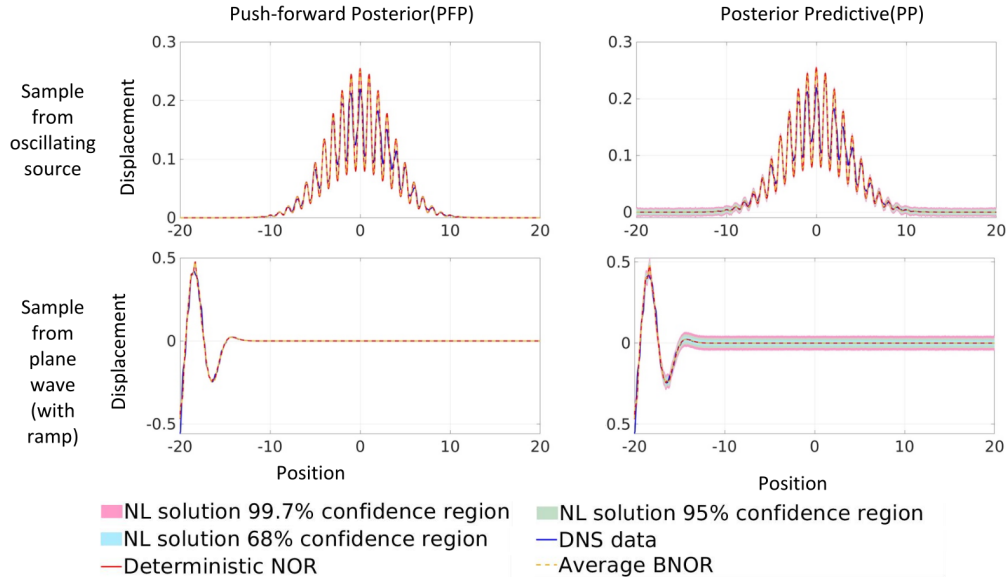


Fig. 11. Comparison of two posterior uncertainty results for the prediction of two exemplar examples from the training dataset, on the random microstructure case. Left: displacement field prediction and confidence regions from the push-forward posterior (PFP) approach. Right: displacement field prediction and confidence regions from the posterior predictive (PP) approach.

We illustrate again, in Fig. 11, posterior uncertainty on two predicted solutions that are part of the random microstructure training data, including both the PFP and PP posterior uncertainty results. We can see that the BNOR solution with the mean $\mathbf{C}$, the deterministic NOR, and the DNS solution are all very well matched for the plane wave case. On the other hand, again, the oscillating source case exhibits discernible displacement magnitude differences between the two nonlocal solutions and the DNS solution. We also see that the PFP results in negligible predictive uncertainty, as already observed in Fig. 10. On the other hand, the PP results exhibit non-negligible uncertainty at low mean-output levels for both the oscillating source and plane wave case, while exhibiting again negligible uncertainty at high displacement levels even where the discrepancy with the DNS data is large. Here again, the PP with the current data model does not span the discrepancy from the data well, and is a candidate for improvement in future work.

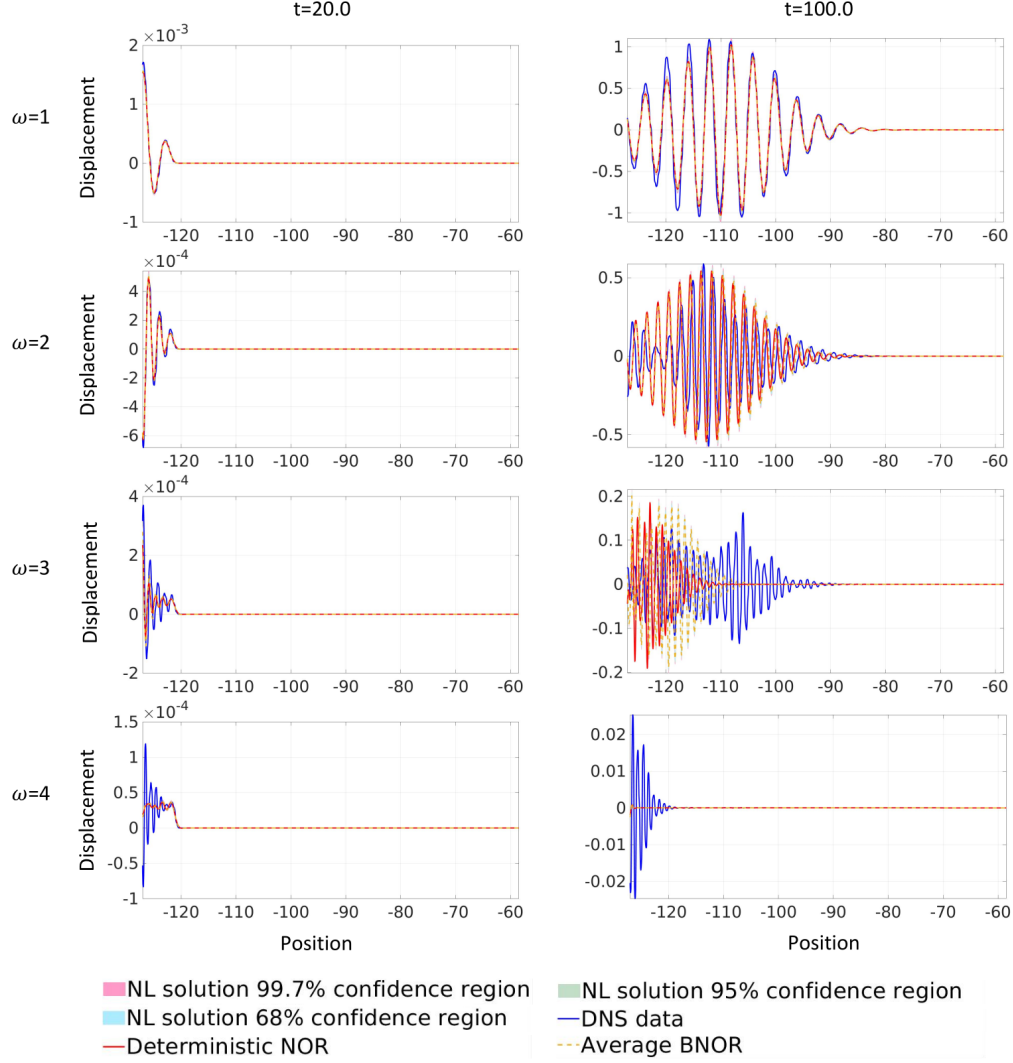


Fig. 12. Validation on the wave packet traveling problem for a bar with random microstructure, illustrating PFP uncertainty. Plots from top to bottom are corresponding to different loading frequencies: 1) $\omega = 1 < \omega_{bs}$, 2) $\omega = 2 < \omega_{bs}$, 3) $\omega = 3 \approx \omega_{bs}$, and 4) $\omega = 4 > \omega_{bs}$. Left column shows the simulation results on a relatively short time ($T = 20$), and the right column demonstrates long time simulation results ($T = 100$).

Validation on wave packet Following the same procedure as for the periodic microstructure case, here we again demonstrate the generalization capability of the learnt nonlocal surrogate model on different domains, boundary conditions, and longer simulation time, by considering the wave packet problem. Prediction results subject to four loading frequencies, $\omega = 1.0 < \omega_{bs}$, $\omega = 2.0 < \omega_{bs}$, $\omega = 3.0 \approx \omega_{bs}$, and $\omega = 4.0 > \omega_{bs}$, are provided on a relatively short time ($t = 20$) and a longer time ($t = 100$) simulations. The validation results are illustrated in Figures 12 and 13 illustrating

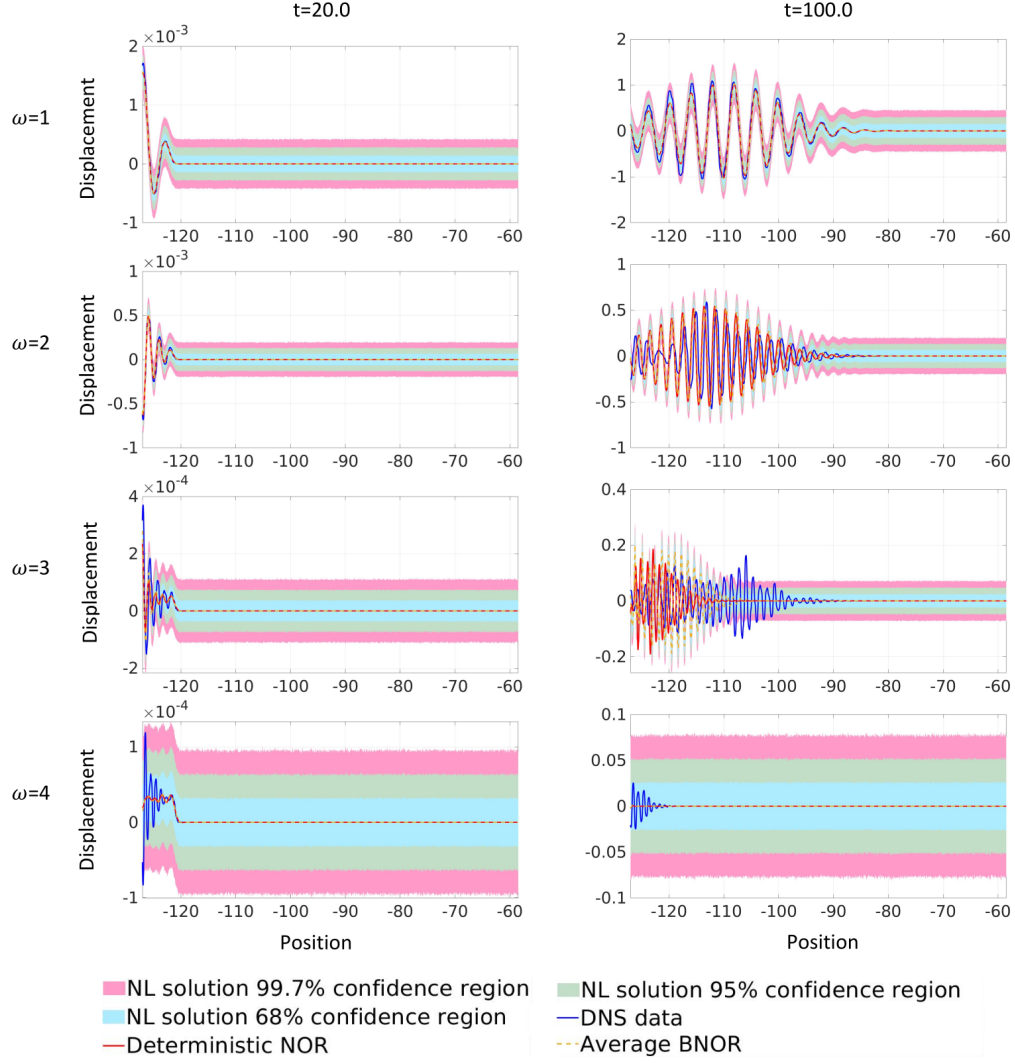


Fig. 13. Validation on the wave packet traveling problem for a bar with random microstructure, illustrating PP uncertainty. Plots from top to bottom are corresponding to different loading frequencies: 1) $\omega = 1 < \omega_{bs}$, 2) $\omega = 2 < \omega_{bs}$, 3) $\omega = 3 \approx \omega_{bs}$, and 4) $\omega = 4 > \omega_{bs}$. Left column shows the simulation results on a relatively short time ($T = 20$), and the right column demonstrates long time simulation results ($T = 100$).

PFP and PP uncertainty respectively. The deterministic nonlocal solution from the original NOR algorithm and the DNS data are also reported for comparison.

We can see here again the negligible PFP uncertainty, even under extrapolation, in Fig. 12, while significant PP uncertainty is evident in Fig. 13. Further, for the short time prediction task ($t = 20$), a good agreement between the mean of the uncertain nonlocal solution and the DNS data is observed for $\omega = 1, 2$, validating the generalizability of the learnt nonlocal surrogate under these

loading conditions up to this time. However, there is clear mismatch with DNS for $\omega = 3, 4$ in the region near the wave front. For long time prediction ($t = 100$), the extrapolated learnt nonlocal surrogate model exhibits discernible differences from the DNS solution under all loading cases, with the minimal differences being evident at $\omega = 1$. At this loading the differences are confined to small magnitude differences in the oscillatory trace. On the other hand, more significant differences in both magnitude and phase are evident for $\omega = 2, 3, 4$. Qualitatively, a satisfactory result is indeed obtained when the loading frequency $\omega = 4.0 > \omega_{bs}$ in that the wave barely travels in time. Except for the case with $t = 100$ and $\omega = 3$, the DNS solution always lies inside the 68% PP confidence region, highlighting the efficacy of our additive error model under these conditions in extrapolation.

CONCLUSION

In this paper, we have proposed BNOR – an approach for learning the optimal nonlocal surrogate together with modeling discrepancy characterization for heterogeneous material homogenization. Our work is built based on the nonlocal operator regression approach (You et al. 2021b), which aims to provide a well-posed and generalizable nonlocal surrogate model from high-fidelity simulations and/or experimental measurements of the displacement fields, and allows for accurate simulations at a larger scale than the microstructure. Because of these desired properties, the nonlocal surrogate model from NOR is readily applicable for unseen prediction tasks, such as to find the solutions at much larger times than the time instants used for training, and on problem settings that are substantially different from the training data set. On the other hand, since the nonlocal surrogate model from NOR serves as a homogenized surrogate of the original complex physical system without resolving its heterogeneities at the microscale, unavoidable modeling discrepancy will be introduced, which contribute to the overall prediction error and uncertainty.

To quantify the model predictions and associated uncertainty in future prediction tasks, in this work we used an independent additive Gaussian data model centered on the NOR approach, in order to represent and quantify the homogenization modeling discrepancies. The framework is developed within a Bayesian inference context, where NOR model parameters are inferred simultaneously along with parameters that characterize errors relative to the data. To solve the Bayesian

inference problem efficiently, a two-phase MCMC algorithm is proposed, enabling an efficient and non-intrusive procedure for approximate likelihood construction and model discrepancy estimation. Lastly, the proposed BNOR framework is validated on the wave propagation problem in heterogeneous bars. It is found that the learnt model has 1) captured the correct band gap and the group velocity; 2) reproduced high-fidelity data for a composite material in applications that are substantially different from the training data under a significant range of conditions; 3) provided a characterization for the posterior distribution of the parameters as well as the confidence region for further prediction tasks.

In this work, we have focused on the high-fidelity measurements so the modeling discrepancy acts as the main contributor of uncertainties. As observed in the numerical tests of Section “Application to a Heterogeneous Elastic Bar”, the posterior predictive does not span the discrepancy in the data well, due to the limitation of our additive *i.i.d.* Gaussian error Likelihood construction. As a natural follow-up, and to provide predictive uncertainty that better matches the point-wise discrepancies between the calibrated model and the data, we plan to pursue a more sophisticated embedded model error construction (Sargsyan et al. 2015; Sargsyan et al. 2019). Another interesting future direction would be to incorporate our model while enabling targeted model improvement and optimal experimental design. Last but not least, to illustrate the efficiency of our algorithm, two- and three-dimensional test cases will also be considered. For example, one may use the BNOR framework to characterize the homogenization error in the coarse-grained nonlocal model from molecular dynamics simulations, as an extension of the development in (You et al. 2022).

Data Availability Statement

All data, models, or code that support the findings of this study are available from the corresponding author upon reasonable request.

Acknowledgments

Y. Fan and Y. Yu would like to acknowledge support by the National Science Foundation under award DMS-1753031 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.

S. Silling and M. D’Elia would like to acknowledge the support of the Sandia National Laboratories (SNL) Laboratory-directed Research and Development program and by the U.S. Department of Energy (DOE), Office of Advanced Scientific Computing Research (ASCR) under the Collaboratory on Mathematics and Physics-Informed Learning Machines for Multiscale and Multiphysics Problems (PhILMs) project. H. Najm acknowledges support by the U.S. DOE ASCR office, Scientific Discovery through Advanced Computing (SciDAC) program. This article has been authored by an employee of National Technology and Engineering Solutions of Sandia, LLC under Contract No. DE-NA0003525 with the U.S. Department of Energy (DOE). The employee owns all right, title and interest in and to the article and is solely responsible for its contents. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this article or allow others to do so, for United States Government purposes. The DOE will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan <https://www.energy.gov/downloads/doe-public-access-plan>.

REFERENCES

- Andrieu, C., De Freitas, N., Doucet, A., and Jordan, M. I. (2003). “An introduction to mcmc for machine learning.” *Machine learning*, 50(1), 5–43.
- Andrieu, C. and Thoms, J. (2008). “A tutorial on adaptive mcmc.” *Statistics and computing*, 18(4), 343–373.
- Bensoussan, A., Lions, J.-L., and Papanicolaou, G. (2011). *Asymptotic analysis for periodic structures*, Vol. 374. American Mathematical Soc.
- Beran, M. and McCoy, J. (1970). “Mean field variations in a statistical sample of heterogeneous linearly elastic solids.” *International Journal of Solids and Structures*, 6(8), 1035–1054.
- Bobaru, F., Foster, J. T., Geubelle, P. H., and Silling, S. A. (2016). *Handbook of peridynamic modeling*. CRC press.

- Carlin, B. P. and Louis, T. A. (2011). *Bayesian Methods for Data Analysis*. Chapman and Hall/CRC, Boca Raton, FL.
- Casella, G. and Berger, R. L. (1990). *Statistical inference*, Vol. 70. Duxbury Press Belmont, CA.
- Cherednichenko, K., Smyshlyaev, V. P., and Zhikov, V. (2006). “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.
- Cui, T., Marzouk, Y., and Willcox, K. (2016). “Scalable posterior approximations for large-scale Bayesian inverse problems via likelihood-informed parameter and state reduction.” *Journal of Computational Physics*, 315, 363–387.
- de Moraes, E. A. B., D’Elia, M., and Zayernouri, M. (2022). “Nonlocal machine learning of micro-structural defect evolutions in crystalline materials.” *arXiv preprint arXiv:2205.05729*.
- Debusschere, B., Najm, H., Pébay, P., Knio, O., Ghanem, R., and Le Maître, O. (2004). “Numerical challenges in the use of polynomial chaos representations for stochastic processes.” *SIAM Journal on Scientific Computing*, 26(2), 698–719.
- Debusschere, B., Sargsyan, K., Safta, C., and Chowdhary, K. (2017). “The uncertainty quantification toolkit (uqtk).” *Handbook of Uncertainty Quantification*, R. Ghanem, D. Higdon, and H. Owhadi, eds., Springer, 1807–1827.
- Dobson, M., Luskin, M., and Ortner, C. (2010). “Sharp stability estimates for the force-based quasicontinuum approximation of homogeneous tensile deformation.” *Multiscale Modeling & Simulation*, 8(3), 782–802.
- Du, Q., Engquist, B., and Tian, X. (2020). “Multiscale modeling, homogenization and nonlocal effects: Mathematical and computational issues.” *Contemporary mathematics*, 754.
- Du, Q., Gunzburger, M., Lehoucq, R. B., and Zhou, K. (2013). “A nonlocal vector calculus, nonlocal volume-constrained problems, and nonlocal balance laws.” *Mathematical Models and Methods in Applied Sciences*, 23(03), 493–540.
- Du, Q., Tao, Y., and Tian, X. (2017). “A peridynamic model of fracture mechanics with bond-breaking.” *Journal of Elasticity*, 1–22.

- Du, Q. and Zhou, K. (2011). “Mathematical analysis for the peridynamic nonlocal continuum theory.” *ESAIM: Mathematical Modelling and Numerical Analysis*, 45(02), 217–234.
- Efendiev, Y., Galvis, J., and Hou, T. Y. (2013). “Generalized multiscale finite element methods (gmsfem).” *Journal of computational physics*, 251, 116–135.
- Eringen, A. C. and Edelen, D. G. B. (1972). “On nonlocal elasticity.” *International Journal of Engineering Science*, 10(3), 233–248.
- Haario, H., Saksman, E., and Tamminen, J. (2001). “An adaptive Metropolis algorithm.” *Bernoulli*, 7, 223–242.
- Hakim, L., Lacaze, G., Khalil, M., Sargsyan, K., Najm, H., and Oefelein, J. (2018). “Probabilistic parameter estimation in a 2-step chemical kinetics model for n-dodecane jet autoignition.” *Combustion Theory and Modeling*, 22(3), 446–466.
- Hansen, P. C. (2000). *The L-Curve and its Use in the Numerical Treatment of Inverse Problems*. WIT Press.
- Higdon, D., Lee, H., and Holloman, C. (2003). “Markov chain Monte Carlo-based approaches for inference in computationally intensive inverse problems.” *Bayesian Statistics*, 7, 181–197.
- Huan, X., Safta, C., Sargsyan, K., Geraci, G., Eldred, M. S., Vane, Z. P., Lacaze, G., Oefelein, J. C., and Najm, H. N. (2018). “Global Sensitivity Analysis and Estimation of Model Error, toward Uncertainty Quantification in Scramjet Computations.” *AIAA Journal*, 56(3), 1170–1184.
- Hughes, T. J., Wells, G. N., and Wray, A. A. (2004). “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.
- Jaynes, E. (2003). *Probability Theory: The Logic of Science*, G.L. Bretthorst, Ed. Cambridge University Press, Cambridge, UK.
- Junghans, C., Praprotnik, M., and Kremer, K. (2008). “Transport properties controlled by a thermostat: An extended dissipative particle dynamics thermostat.” *Soft Matter*, 4(1), 156–161.
- Karal Jr, F. C. and Keller, J. B. (1964). “Elastic, electromagnetic, and other waves in a random

medium.” *Journal of Mathematical Physics*, 5(4), 537–547.

Kennedy, M. and O’Hagan, A. (2000). “Predicting the output from a complex computer code when fast approximations are available.” *Biometrika*, 87(1), 1–13.

Kennedy, M. C. and O’Hagan, A. (2001). “Bayesian calibration of computer models.” *Journal of the Royal Statistical Society: Series B*, 63(3), 425–464.

Kubo, R. (1966). “The fluctuation-dissipation theorem.” *Reports on progress in physics*, 29(1), 255.

Lang, Q. and Lu, F. (2022). “Learning interaction kernels in mean-field equations of first-order systems of interacting particles.” *SIAM Journal on Scientific Computing*, 44(1), A260–A285.

Laplace, M. d. (1814). *Essai philosophique sur les probabilités*. Gauthier-Villars, Paris English Ed.: "A Philosophical Essay on Probabilities", Dover Pub., 6th ed., 1995.

Lu, F., An, Q., and Yu, Y. (2022). “Nonparametric learning of kernels in nonlocal operators.” *arXiv preprint arXiv:2205.11006*.

Moës, N., Oden, J. T., Vemaganti, K., and Remacle, J.-F. (1999). “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.

Oliver, T. A. and Moser, R. D. (2011). “Bayesian uncertainty quantification applied to RANS turbulence models.” *J. Phys.: Conf. Ser.*, 318.

Ortiz, M. (1987). “A method of homogenization of elastic media.” *International journal of engineering science*, 25(7), 923–934.

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

Robert, C. and Casella, G. (2004). *Monte Carlo Statistical Methods*. Springer Texts in Statistics. Springer.

Santosa, F. and Symes, W. W. (1991). “A dispersive effective medium for wave propagation in periodic composites.” *SIAM Journal on Applied Mathematics*, 51(4), 984–1005.

- Sargsyan, K., Huan, X., and Najm, H. (2019). “Embedded model error representation for Bayesian model calibration.” *Int. J. UQ*, 9(4), 365–394.
- Sargsyan, K., Najm, H., and Ghanem, R. (2015). “On the Statistical Calibration of Physical Models.” *International Journal of Chemical Kinetics*, 47(4), 246–276.
- Silling, S. A. (2000). “Reformulation of elasticity theory for discontinuities and long-range forces.” *Journal of the Mechanics and Physics of Solids*, 48(1), 175–209.
- Silling, S. A. (2021). “Propagation of a stress pulse in a heterogeneous elastic bar.” *Journal of Peridynamics and Nonlocal Modeling*, 3(3), 255–275.
- Silling, S. A., D’Elia, M., Yu, Y., You, H., and Fermen-Coker, M. (2022). “Peridynamic model for single-layer graphene obtained from coarse-grained bond forces.” *Journal of Peridynamics and Nonlocal Modeling*, 1–22.
- Sivia, D. S. and Skilling, J. (2006). *Data Analysis: A Bayesian Tutorial, Second Edition*. Oxford University Press.
- Smyshlyaev, V. P. and Cherednichenko, K. D. (2000). “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.
- Suzuki, J., Gulian, M., Zayernouri, M., and D’Elia, M. (2022). “Fractional modeling in action: A survey of nonlocal models for subsurface transport, turbulent flows, and anomalous materials.” *Journal of Peridynamics and Nonlocal Modeling*.
- Tauchert, T. R. and Guzelsu, A. (1972). “An experimental study of dispersion of stress waves in a fiber-reinforced composite.” *Journal of Applied Mechanics*, 39(1), 98–102.
- Trillos, N. G. and Sanz-Alonso, D. (2017). “The bayesian formulation and well-posedness of fractional elliptic inverse problems.” *Inverse Problems*, 33(6), 065006.
- Vats, D., Flegal, J. M., and Jones, G. L. (2019). “Multivariate output analysis for markov chain monte carlo.” *Biometrika*, 106(2), 321–337.
- Weinan, E. and Engquist, B. (2003). “Multiscale modeling and computation.” *Notices of the AMS*, 50(9), 1062–1070.

- Willis, J. R. (1985). “The nonlocal influence of density variations in a composite.” *International Journal of Solids and Structures*, 21(7), 805–817.
- Xu, X., D’Elia, M., Glusa, C., and Foster, J. T. (2022). “Machine-learning of nonlocal kernels for anomalous subsurface transport from breakthrough curves.” *NACO*.
- Xu, X., D’Elia, M., and Foster, J. T. (2021). “A machine-learning framework for peridynamic material models with physical constraints.” *Computer Methods in Applied Mechanics and Engineering*, 386, 114062.
- Xu, X. and Foster, J. T. (2020). “Deriving peridynamic influence functions for one-dimensional elastic materials with periodic microstructure.” *Journal of Peridynamics and Nonlocal Modeling*, 2(4), 337–351.
- You, H., Yu, Y., Silling, S., and D’Elia, M. (2021a). “Data-driven learning of nonlocal models: from high-fidelity simulations to constitutive laws.” *AAAI Spring Symposium: MLPS*.
- You, H., Yu, Y., Silling, S., and D’Elia, M. (2022). “A data-driven peridynamic continuum model for upscaling molecular dynamics.” *Computer Methods in Applied Mechanics and Engineering*, 389, 114400.
- You, H., Yu, Y., Trask, N., Gulian, M., and D’Elia, M. (2021b). “Data-driven learning of non-local physics from high-fidelity synthetic data.” *Computer Methods in Applied Mechanics and Engineering*, 374, 113553.
- Zhang, L., You, H., and Yu, Y. (2022). “MetaNOR: A meta-learned nonlocal operator regression approach for metamaterial modeling.” *MRS Communications*.
- Zohdi, T. I. (2017). “Homogenization methods and multiscale modeling.” *Encyclopedia of Computational Mechanics Second Edition*, 1–24.

APPENDIX: ABLATION STUDY

Herein, we provide additional numerical results to investigate the effect of the proposed initialization scheme in phase 1. In particular, we perform the phase 2 in our algorithm by starting from an all-one initial state. As shown in the trace plots of Figure 14, the acceptance rate is very low (0.0059) and only a few proposals are sampled.

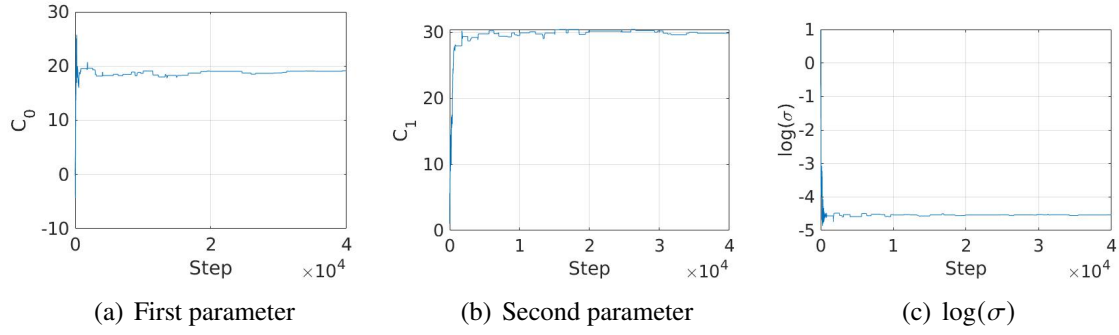


Fig. 14. Ablation study – Skip phase 1 and run phase 2 directly: The first two kernel parameters C_0 , C_1 and $\log(\sigma)$ for a bar with periodic microstructure, plotted against the iteration number (the trace plot of MCMC).