

Hetero-EUCLID: Interpretable model discovery for heterogeneous hyperelastic materials using stress-unsupervised learning

Kanhaiya Lal Chaurasiya^a, Saurav Dutta^a, Siddhant Kumar^b, Akshay Joshi^{a,*}

^a*Department of Mechanical Engineering, Indian Institute of Science, Bengaluru, India*

^b*Department of Materials Science and Engineering, Delft University of Technology, 2628 CD Delft, The Netherlands*

Abstract

We propose a computational framework, Hetero-EUCLID, for segmentation and parameter identification to characterize the full hyperelastic behavior of all constituents of a heterogeneous material. In this work, we leverage the Bayesian-EUCLID (Efficient Unsupervised Constitutive Law Identification and Discovery) framework to efficiently solve the heterogenized formulation through parsimonious model selection using sparsity-promoting priors and Monte Carlo Markov Chain sampling. We utilize experimentally observable 3D surface displacement and boundary-averaged force data generated from Finite Element simulations of non-equibiaxial tension tests on heterogeneous specimens. The framework broadly consists of two steps—residual force-based segmentation, and constitutive parameter identification. We validate and demonstrate the ability of the proposed framework to segment the domain, and characterize the constituent materials on various types of thin square heterogeneous domains. We validate of the framework's ability to segment and characterize materials with various levels of displacement noises and non-native mesh discretizations, i.e., using different meshes for the forward FE simulations and the inverse EUCLID problem. This demonstrates Hetero-EUCLID framework's applicability in Digital Image/Volume Correlation-based experimental scenarios. Furthermore, the proposed framework performs successful segmentation and material characterizations based on data from a single experiment, thereby making it viable for rapid, interpretable model discovery in domains such as aerospace and defense composites and for characterization of selective tissue stiffening in medical conditions such as *fibroatheroma*, *atherosclerosis*, or cancer.

Keywords: Heterogeneous material, data-driven discovery, hyperelastic constitutive models, Bayesian learning, Uncertainty quantification

1. Introduction

With the increasing relevance of biomedical engineering, there is a growing focus on understanding the mechanical behavior of biological tissues. Many of these tissues exhibit intrinsic material heterogeneity [1–3], necessitating the development of multi-segmented constitutive models. Such models can significantly enhance our understanding of complex biomechanical processes, including the progression of aortic aneurysms [4–6] and the localized stiffening of tissues due to tumor growth [7, 8].

A more traditional approach to constitutive modeling had been to perform the phenomenological model calibration of an *apriori* assumed constitutive equation, utilizing the full-field displacement data obtained from Digital Image Correlation (DIC) in combination with boundary-averaged force measurement from force transducers [9]. The inverse problem of determining the constitutive parameters is solved using Finite Element Model Updating (FEMU) [10] or Virtual Fields Method (VFM) [11]. However, assuming a form of the constitutive equation for the material limits the materials and constitutive models that can be characterized

*Corresponding author

Email address: akshayjoshi@iisc.ac.in (Akshay Joshi)

with this approach. To address the limitations of traditional model calibration, and enable characterization and discovery of materials that are not necessarily well explained by existing models, several works in the past have sought to entirely bypass [12–19] or surrogate constitutive models. The approach of surrogating constitutive model identifies a relationship between mechanical stresses and strains using Gaussian Process Regression (GPR) [20, 21], Radial Basis Functions (RBF) [22], piece-wise polynomial interpolation [23, 24], and Artificial Neural Networks (ANN) [25–35]. Furthermore, to address the lack of availability of full stress-tensor data for training supervised training several frameworks leverage the balance of linear momentum [36] or the virtual fields method [37] to train ANNs or Recurrent Neural Networks (RNNs) to surrogate the stress-strain relationship for the materials using only experimentally available surface displacement and boundary-averaged force data.

Although stress-unsupervised ANN/RNN surrogates are capable of learning complicated constitutive models, they do not provide interpretable constitutive models. The EUCLID framework [38] was initially introduced to model complicated homogeneous hyperelastic constitutive behavior in an interpretable manner. This was achieved through sparse symbolic regression using a pre-defined library of known strain energy densities. The EUCLID framework has since been extended to model homogeneous elastoplastic materials [39], viscoelastic materials [40], and generalized standard materials [41]. Most of the aforementioned works address the modeling of homogeneous materials.

An earlier work which used only the surface displacement data [42] focused on elastography of incompressible linear elastic heterogeneous solids using novel Physics Informed Neural Networks (PINNs) to predict the variation of the Young's Modulus as a function of position in the material. A later work by the same authors employed multiple PINNs to obtain the spatial variation of Poisson's ratio in elastography [43]. Another work presented a detailed study on segmenting heterogeneous hyperelastic materials using surface displacement data and associated strain invariants [44]. A more recent work used a stress-unsupervised ANN-based framework [45] to segment and model a heterogeneous hyperelastic solid, wherein the inputs to the ANN are the strain invariants and the query position, and the output of the ANN is the strain energy density. The novel ANN-based framework was able to segment and reproduce the constitutive models of each segment in the hyperelastic solid. However, the use of ANNs makes the discovered models uninterpretable and provides no uncertainty estimates on the discovered models. Furthermore, since the query position is an essential input to the aforementioned ANN, it would complicate the ANN's adoption as a constitutive model for forward simulation on other geometries. In this context, we propose the Hetero-EUCLID framework, wherein we extend the Bayesian-EUCLID framework [46] to segment a heterogeneous solid and provide interpretable constitutive models for all constituents, using only surface displacement and boundary-averaged force data.

The problem of constitutively modeling all the constituents in a heterogeneous geometry can be divided into two subtasks: segmenting the geometry and obtaining the constitutive model for all the materials involved. In [section 2](#) of the manuscript we present

- An overview of extending the Bayesian-EUCLID framework [46], which was originally developed for homogeneous materials, to model segmented homogeneous materials.
- A survey of existing segmenting techniques that rely only on surface displacement data, and the reason for our choice of the residual force method.
- A comment on displacement-noise and the steps taken to mitigate its effect on the proposed framework's results.

A schematic illustration of the proposed Hetero-EUCLID framework is provided in [Figure 1](#). In [section 3](#), we present results of validation studies on square geometries containing two or more hyperelastic constituents. The validation studies consist of using surface displacement data and boundary force data from biaxial tension tests on thin hyperelastic specimens. For the scope of this study, we source displacement and boundary-averaged force data from 3D Finite Element Simulations. This is in contrast to previous similar studies, which focused on validating their results on 2D plane strain settings [36, 45]. In [section 3](#) we also validate the proposed Hetero-EUCLID framework on "Non-native" meshed geometries, wherein the Finite Element mesh used to perform the inverse Hetero-EUCLID analysis was different from the Finite Element

mesh used to generate the displacement and boundary-averaged force. The manuscript concludes in section 4 with comments on the results of the study and potential future work.

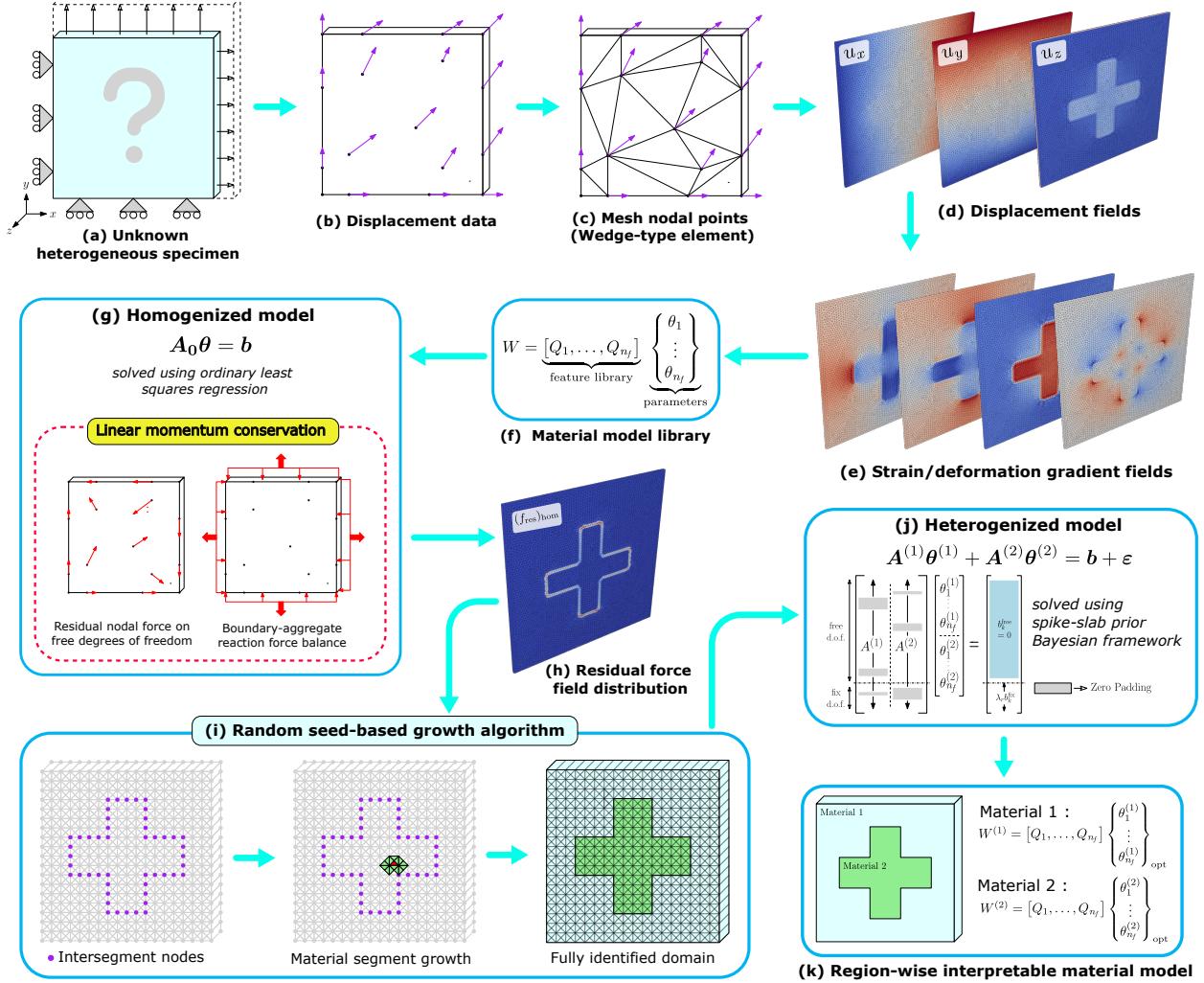


Figure 1: Schematic of the Hetero-EUCLID workflow for unsupervised constitutive model discovery in heterogeneous hyperelastic materials. (a) illustrates the specimen subjected to non-equi-biaxial tensile loading. (b) and (c) show the point-wise displacement observations and corresponding finite element interpolation onto a wedge-type mesh, respectively. (d) depicts the resulting full-field displacement field, while (e) presents the corresponding strain field obtained via spatial differentiation. (f) represents a catalogue of the hyperelastic constitutive models library. (g) outlines the homogenized model formulation incorporating the weak form of the linear momentum balance, which is solved using ordinary least squares regression to yield the residual force norm distribution shown in (h). Material segmentation based on the residual force indicator is performed using a random seed-based island growth algorithm, as illustrated in (i). (j) presents the unified global heterogenized system of equations assembled after material segmentation. The unknown constitutive models $\theta^{(i)}$ are solved for using the Bayesian EUCLID framework. (k) highlights the discovered region-wise interpretable constitutive models corresponding to each material segment within the heterogeneous specimen.

2. Hetero-EUCLID: Methodology

2.1. Incorporating heterogeneity

Consider a three-dimensional reference domain $\Omega \in \mathbb{R}^3$, with boundary $\partial\Omega$ in a plane stress condition subjected to quasi-static displacement-controlled non-equi-biaxial loading on $\partial\Omega$. Let $\chi = \{\mathbf{X}^a \in \Omega : a =$

$1, \dots, n_n\}$ denote n_n reference points in the given three-dimensional setting with displacement measurements as $\{\mathbf{u}^{a,t} \in \mathbb{R}^3 : a = 1, \dots, n_n; t = 1, \dots, n_t\}$ for n_t time steps. In addition, the boundary-aggregate reaction force measurements at $n_b (\ll n_n)$ load transducer points can be denoted by $\{R_i^{k,t} : k = 1, \dots, n_b; t = 1, \dots, n_t; i = 1, 2, 3\}$.

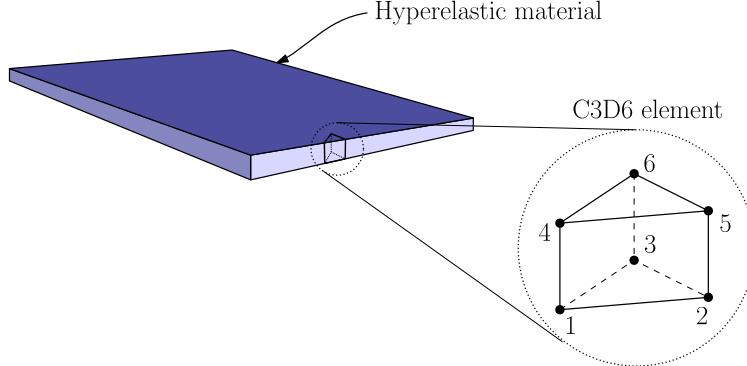


Figure 2: Meshing the given thin plate geometry with C3D6 elements that have a thickness identical to that of the plate.

As shown in Figure 2, the given reference domain, a thin square plate, has been meshed with 6-node linear triangular prism elements (wedge or C3D6) using the nodal set χ . Given nodal displacements $\mathbf{u}^{a,t}$, the value of the displacement field at any point \mathbf{X} in the volume is

$$\mathbf{u}^t(\mathbf{X}) = \sum_{a=1}^{n_n} \mathbb{N}^a(\mathbf{X}) \mathbf{u}^{a,t} \quad (1)$$

where $\mathbb{N}^a : \Omega \rightarrow \mathbb{R}$ denotes the interpolation shape function associated with the a^{th} node in the reference domain. The corresponding deformation gradient can thus be computed as

$$\mathbf{F}^t(\mathbf{X}) = \mathbf{I} + \sum_{a=1}^{n_n} \mathbf{u}^{a,t} \otimes \nabla \mathbb{N}^a(\mathbf{X}) \quad (2)$$

where $\nabla(\cdot) = \partial(\cdot)/\partial \mathbf{X}$ is defined as the gradient operator in the reference coordinates. Due to the use of linear shape functions in the wedge mesh type, $\nabla \mathbb{N}^a(\mathbf{X})$ remains constant within each element. For the scope of this work, a single-element-thick mesh is generated over the reference domain. This would imply that all nodes are located on the surface and their displacements $\mathbf{u}^{a,t}$ could be obtained through DIC experiments. Further, for brevity, in this study we utilize data from a single large deformation, i.e., $t = 1$. We will henceforth be dropping the use of t in the superscript.

We now consider a reference domain composed of multiple hyperelastic material segments. Since each segment belongs to the hyperelastic material class, they can be assigned a strain energy density $W(\mathbf{F})$ and the corresponding first Piola-Kirchhoff stress tensor would be $\mathbf{P}(\mathbf{F}) = \partial W/\partial \mathbf{F}$. In order to satisfy the material objectivity constraint, the strain energy density will be a function of the right Cauchy-Green deformation tensor $\mathbf{C} = \mathbf{F}^T \mathbf{F}$. Furthermore, given that we focus on isotropic materials for this study, the strain energy density $W(\mathbf{C})$ can be defined in terms of isotropic invariants (I_1, I_2, I_3) as

$$W(\mathbf{C}) = W(I_1, I_2, I_3) \quad (3)$$

where,

$$I_1(\mathbf{C}) = \text{tr}(\mathbf{C}), \quad I_2(\mathbf{C}) = \frac{1}{2}[\text{tr}^2(\mathbf{C}) - \text{tr}(\mathbf{C}^2)], \quad I_3(\mathbf{C}) = \det(\mathbf{C}) \quad (4)$$

We address the inverse problem of identifying the strain energy density function for a specimen subjected to a non-equi-biaxial loading test by introducing a material feature library $\mathbf{Q} : \mathbb{R}^3 \rightarrow \mathbb{R}^{n_f}$, where n_f denotes

the number of nonlinear isotropic feature functions. Accordingly, the constitutive relation in [Equation 3](#) can be reformulated using this feature library \mathbf{Q} and a vector of unknown material parameters $\boldsymbol{\theta} \in \mathbb{R}_+^{n_f}$, consisting of non-negative scalar coefficients. Following [46], we impose the constraint on each component of $\boldsymbol{\theta}$ to be non-negative to ensure polyconvexity of $W(I_1, I_2, I_3)$. It should be noted that any terms involving $(\tilde{I}_2 - 3)$ in the feature library \mathbf{Q} are not polyconvex, and should ideally be replaced with $(\tilde{I}_2^{3/2} - 3^{3/2})$ [47]. However, we choose to retain the more traditional $(\tilde{I}_2 - 3)$ term, which appears in the Isihara model [48], and is not known to cause any physical admissibility issues.

$$W(I_1, I_2, I_3; \boldsymbol{\theta}) = \mathbf{Q}^T(I_1, I_2, I_3) \boldsymbol{\theta} \quad (5)$$

For the scope of this work to demonstrate the capability of identifying the heterogeneity in the specimen, we have appropriately chosen the following feature library, which can represent the compressible hyperelastic isotropic material models such as generalized Mooney-Rivlin, which utilizes the isotropic invariants (I_1, I_2, I_3) covering Neo-Hookean [49], Isihara [48], Biderman [50], and Haines-Wilson [51].

$$\begin{aligned} \mathbf{Q}(I_1, I_2, I_3) &= \underbrace{\left[(\tilde{I}_1 - 3)^i (\tilde{I}_2 - 3)^{j-i} : j \in \{1, \dots, N\}, i \in \{0, \dots, j\} \right]^T}_{\text{Generalized Mooney-Rivlin features}} \\ &\oplus \underbrace{\left[(J - 1)^{2k} : k \in \{1, \dots, M\} \right]^T}_{\text{Volumetric energy features}} \end{aligned} \quad (6)$$

where $\tilde{I}_1 = J^{-2/3} I_1$, $\tilde{I}_2 = J^{-4/3} I_2$, $J = \det(\mathbf{F}) = I_3^{1/2}$ are the isotropic invariants of the Cauchy-Green deformation tensor (\mathbf{C}) and the symbol \oplus represents the concatenation of the features of the chosen library. It is important to note that the requirements of satisfying physical constraints such as objectivity, material symmetry, and stress-free reference configuration ($\mathbf{P}(\mathbf{F} = \mathbf{I}) = 0$) are automatically satisfied by [Equation 5](#) and [6](#).

Our approach is based on the stress-unsupervised EUCLID framework, where the learning of material models is performed without relying on experimentally inaccessible stress tensors. To achieve this, we incorporate additional physics-based constraints in the form of conservation of linear momentum. Assuming negligible body forces, the weak form of the linear momentum balance equation under quasi-static loading conditions is applied to the three-dimensional reference domain Ω as follows

$$\int_{\Omega} \mathbf{P} : \nabla \mathbf{v} \, dV - \int_{\partial\Omega} \hat{\mathbf{t}} \cdot \mathbf{v} \, dS = \mathbf{0} \quad \forall \quad \mathbf{v} \text{ differentiable in } \Omega \quad (7)$$

where $\hat{\mathbf{t}}$ represents the traction acting on $\partial\Omega$. Using [Equation 5](#), the first Piola-Kirchhoff stress tensor in indicial notation can be written as

$$P_{iJ} = \frac{\partial W(\mathbf{F})}{\partial F_{iJ}} = \frac{\partial \mathbf{Q}^T(I_1, I_2, I_3)}{\partial F_{iJ}} \boldsymbol{\theta} \quad (8)$$

To compute the partial derivatives of the various features in the material library, $\partial \mathbf{Q}^T / \partial F_{iJ}$, we use the automatic differentiation [52] feature in the torch library of Python. An advantage of using the weak form of the linear momentum balance is that it inherently reduces the sensitivity issues associated with the second-order spatial derivatives present in the strong form [53].

Let $\mathcal{D} = \{(a, i) : a = 1, \dots, n_n; i = 1, 2, 3\}$ represents the set of all nodal d.o.f. (degrees of freedom) for the three-dimensional discretized reference domain. The set \mathcal{D} can be further divided into two subsets, $\mathcal{D}^{\text{free}} \subset \mathcal{D}$ consisting of free d.o.f while $\mathcal{D}^{\text{fix}} = \mathcal{D} \setminus \mathcal{D}^{\text{free}}$ features the fixed d.o.f (via Dirichlet constraints). Using the Bubnov-Galerkin approximation, the test functions are computed as follows

$$v_i(\mathbf{X}) = \sum_{a=1}^{n_n} \mathbb{N}^a(\mathbf{X}) v_i^a \quad (9)$$

Substituting \mathbf{v} from [Equation 9](#) into [Equation 7](#), the weak form then can be written as

$$\sum_{a=1}^{n_n} \left[\int_{\Omega} P_{ij} \mathbb{N}_{,j}^a(\mathbf{X}) v_i^a \, dV - \int_{\partial\Omega} \hat{t}_i \mathbb{N}^a(\mathbf{X}) v_i^a \, dS \right] = 0 \quad (10)$$

Now, we choose various nodal values of the test function v_i^a to generate equations to solve for the constitutive properties of the material.

- (a) $v_i^a = 0 \forall (a, i) \in \mathcal{D}^{\text{fix}} - \text{Homogeneous Dirichlet boundary condition}$

Substituting the value of v_i^a in [Equation 10](#), the modified equation is obtained as

$$\sum_{a=1}^{n_n} \left[v_i^a \int_{\Omega} P_{ij} \mathbb{N}_{,j}^a(\mathbf{X}) \, dV \right] = 0 \quad (11)$$

It is easy to show that [Equation 11](#) holds for all values of v_i^a if and only if (substituting P_{ij} using [Equation 8](#))

$$\int_{\Omega} \left(\frac{\partial \mathbf{Q}^T}{\partial F_{ij}} \boldsymbol{\theta}(\mathbf{X}) \right) \mathbb{N}_{,j}^a(\mathbf{X}) \, dV = 0 \quad \forall (a, i) \in \mathcal{D}^{\text{free}} (= \mathcal{D} \setminus \mathcal{D}^{\text{fix}}) \quad (12)$$

[Equation 12](#) consists of an integral over the entire domain Ω . For a domain consisting of n_c material segments, $\Omega = \cup_{k=1}^{n_c} \Omega^{(k)}$. We can thus rewrite [Equation 12](#) as follows

$$\int_{\Omega^{(1)}} \left(\frac{\partial \mathbf{Q}^T}{\partial F_{ij}} \boldsymbol{\theta}^{(1)} \right) \mathbb{N}_{,j}^a(\mathbf{X}) \, dV + \cdots + \int_{\Omega^{(n_c)}} \left(\frac{\partial \mathbf{Q}^T}{\partial F_{ij}} \boldsymbol{\theta}^{(n_c)} \right) \mathbb{N}_{,j}^a(\mathbf{X}) \, dV = 0 \quad \forall (a, i) \in \mathcal{D}^{\text{free}} \quad (13)$$

Within each integral, the material model $\boldsymbol{\theta}(\mathbf{X}) = \boldsymbol{\theta}^{(k)}$, for $\mathbf{X} \in \Omega^{(k)}$; $k \in \{1, \dots, n_c\}$. Since $\boldsymbol{\theta}^{(k)}$ is constant in each material segment it can be taken out of the integral. [Equation 13](#) can thus be rewritten as

$$\mathbf{A}^{\text{free}(1)} \boldsymbol{\theta}^{(1)} + \mathbf{A}^{\text{free}(2)} \boldsymbol{\theta}^{(2)} + \cdots + \mathbf{A}^{\text{free}(j)} \boldsymbol{\theta}^{(j)} + \cdots + \mathbf{A}^{\text{free}(n_c)} \boldsymbol{\theta}^{(n_c)} = \mathbf{0} \quad (14)$$

where $\mathbf{A}^{\text{free}(k)} \in \mathbb{R}^{|\mathcal{D}^{\text{free}}| \times n_f} \quad \forall k \in \{1, \dots, n_c\}$ and $\mathbf{b}^{\text{free}} \in \mathbb{R}^{|\mathcal{D}^{\text{free}}|}$. [Equation 14](#) can be further written as

$$\mathbf{A}^{\text{free}} \boldsymbol{\theta} = \mathbf{b}^{\text{free}} (= \mathbf{0}) \quad (15)$$

where $\mathbf{A}^{\text{free}} = [\mathbf{A}^{\text{free}(1)} \quad \mathbf{A}^{\text{free}(2)} \dots \mathbf{A}^{\text{free}(n_c)}]$ and $\boldsymbol{\theta} = [\boldsymbol{\theta}^{(1)} \quad \boldsymbol{\theta}^{(2)} \dots \boldsymbol{\theta}^{(n_c)}]^T$

$$(b) \quad v_i^a = \begin{cases} 1 & \forall (a, i) \in \mathcal{D}_k^{\text{fix}} \\ 0 & \text{otherwise} \end{cases} \quad \text{for } k = \{1, \dots, n_b\}$$

where, $\mathcal{D}_k^{\text{fix}} \subset \mathcal{D}^{\text{fix}}$ denotes the set of degree of freedoms belonging to the k^{th} boundary $\partial\Omega_k$ and n_b denotes the total number of boundaries where force is measured. Substituting the value of chosen test function v_i^a , [Equation 10](#) gets modified as follows

$$\int_{\Omega_k^e} P_{ij} \mathbb{N}_{,j}^a(\mathbf{X}) \, dV - \int_{\partial\Omega_k} \hat{t}_i \mathbb{N}^a(\mathbf{X}) \, dS = 0 \quad (16)$$

Where the term $\int_{\partial\Omega_k} \hat{t}_i \mathbb{N}^a(\mathbf{X}) dS$ can be interpreted as the total force experimentally recorded on boundary k along the i direction, denoted by R_i^k . Ω_k^e represents volume enclosed by the set of mesh elements e for which at least one node lies on the external boundary $\partial\Omega_k$ of the reference domain Ω . It is important to note that Ω_k^e is different from $\Omega^{(k)}$ defined earlier, wherein the latter corresponds to the volume enclosed by the k^{th} material segment. Further, [Equation 16](#) can be recast as

$$\int_{\Omega_k^e} P_{iJ} \mathbb{N}_{,J}^a(\mathbf{X}) dV = R_i^k \quad (17)$$

Usually, in experiments, load cells measure only one component of force, i.e., normal force. In this case, only component $i \in \{1, 2, 3\}$ of R_i^k would be recorded at a boundary k . However, specialized shear load cells can also be used to measure the other two components. In this study, for convenience, we assume all three components of \mathbf{R}^k are known at every boundary $k \in \{1 \dots n_b\}$. Thus, using [Equation 8](#) and the methodology adopted in Equations [13-15](#), [Equation 17](#) can be written as

$$\begin{pmatrix} \int_{\Omega_1^{e(1)}} \frac{\partial \mathbf{Q}^T}{\partial F_{iJ}} \mathbb{N}_{,J}^a(\mathbf{X}) dV & \dots & \int_{\Omega_1^{e(n_c)}} \frac{\partial \mathbf{Q}^T}{\partial F_{iJ}} \mathbb{N}_{,J}^a(\mathbf{X}) dV \\ \vdots & & \vdots \\ \int_{\Omega_k^{e(1)}} \frac{\partial \mathbf{Q}^T}{\partial F_{iJ}} \mathbb{N}_{,J}^a(\mathbf{X}) dV & \dots & \int_{\Omega_k^{e(n_c)}} \frac{\partial \mathbf{Q}^T}{\partial F_{iJ}} \mathbb{N}_{,J}^a(\mathbf{X}) dV \\ \vdots & & \vdots \\ \int_{\Omega_{n_b}^{e(1)}} \frac{\partial \mathbf{Q}^T}{\partial F_{iJ}} \mathbb{N}_{,J}^a(\mathbf{X}) dV & \dots & \int_{\Omega_{n_b}^{e(n_c)}} \frac{\partial \mathbf{Q}^T}{\partial F_{iJ}} \mathbb{N}_{,J}^a(\mathbf{X}) dV \end{pmatrix} \boldsymbol{\theta} = \begin{pmatrix} R_i^1 \\ \vdots \\ R_i^k \\ \vdots \\ R_i^{n_b} \end{pmatrix} \quad \forall \quad i = 1, 2, 3 \quad (18)$$

where $\Omega_k^{e(j)}$ denotes the volume enclosed by elements of material segment j in the volume Ω_k^e , and $\boldsymbol{\theta}$ retains its definition from [Equation 15](#). [Equation 18](#) results in a system of $3n_b$ linear equations following the numerical integration and thus can be assembled as follows

$$\mathbf{A}^{\text{fix}} \boldsymbol{\theta} = \mathbf{b}^{\text{fix}} \quad \text{with} \quad \mathbf{A}^{\text{fix}} \in \mathbb{R}^{3n_b \times n_f n_c} \quad \text{and} \quad \mathbf{b}^{\text{fix}} \in \mathbb{R}^{3n_b} \quad (19)$$

By combining [Equation 15](#) and [19](#), we formulate a unified model of the form $\mathbf{A}\boldsymbol{\theta} = \mathbf{b}$, where a hyper-parameter λ_r is introduced to control the relative weights between the free and fixed degrees of freedom in the reference domain within the force balance equation.

$$\mathbf{A}\boldsymbol{\theta} = \mathbf{b} \quad \text{with} \quad \mathbf{A} = \begin{bmatrix} \mathbf{A}^{\text{free}} \\ \lambda_r \mathbf{A}^{\text{fix}} \end{bmatrix} \quad \text{and} \quad \mathbf{b} = \begin{bmatrix} \mathbf{b}^{\text{free}} \\ \lambda_r \mathbf{b}^{\text{fix}} \end{bmatrix} \quad (20)$$

[Figure 3](#) provides a schematic to aid in better understanding the compact [Equation 20](#). The global matrix \mathbf{A} is assembled such that its rows contain non-zero entries corresponding to $\mathbf{A}^{(j)}$ only if the associated degrees of freedom belong to material segment j , as illustrated in [Figure 3](#).

To ensure dimensional compatibility, zero-padding is applied to the rows (d.o.f.) of $\mathbf{A}^{(j)}$ not belonging to the material segment j . The resulting right-hand side vector \mathbf{b} consists of two components: \mathbf{b}^{free} , representing the zero internal force condition at nodes in $\mathcal{D}^{\text{free}}$, and \mathbf{b}^{fix} , denoting the boundary-aggregated reaction forces R_i^k ($k \in \{1, \dots, n_b\}; i \in \{1, 2, 3\}$) for the subset \mathcal{D}^{fix} .

Thus, as shown in [Figure 3](#), the problem of obtaining each segment's material parameters $\boldsymbol{\theta}^{(j)}$, can be recast into the form $\mathbf{A}\boldsymbol{\theta} = \mathbf{b}$. From the shape of \mathbf{A} , it can be noted that $|\mathcal{D}^{\text{free}}| + 3n_b \gg n_f n_c$. This would imply $\mathbf{A}\boldsymbol{\theta} = \mathbf{b}$ is an overdetermined system, and the solution for $\boldsymbol{\theta}$ can be obtained in an ordinary least-square sense as

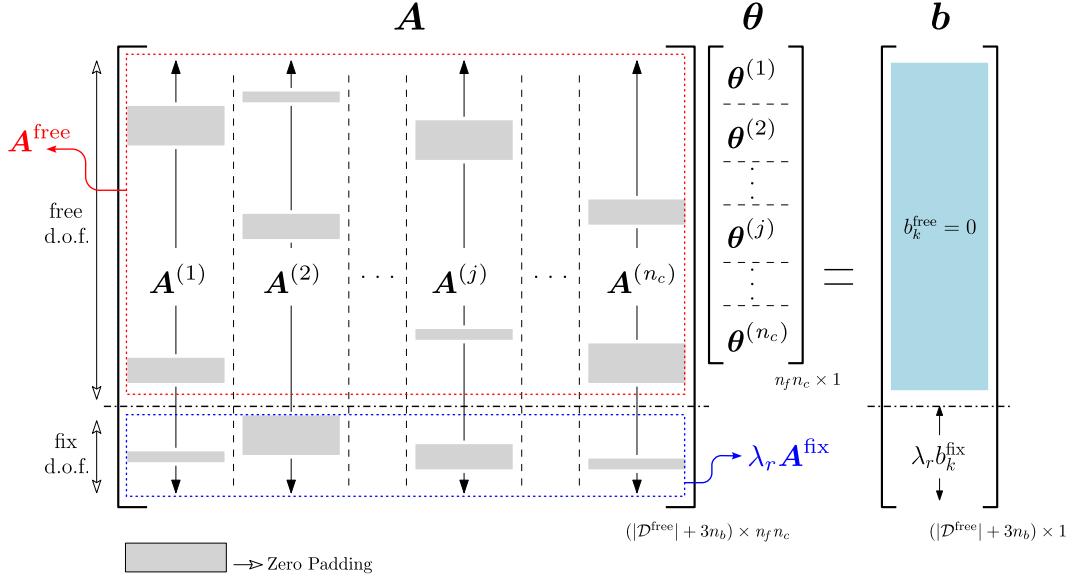


Figure 3: Graphical illustration of the unified global heterogenized model $\mathbf{A}\boldsymbol{\theta} = \mathbf{b}$. The matrix \mathbf{A} comprises $\mathcal{D}^{\text{free}}$ rows corresponding to the free degrees of freedom, along with additional rows for the Dirichlet boundary nodes \mathcal{D}^{fix} . Each segment-specific block $\mathbf{A}^{(j)}$, containing n_f columns associated with the constitutive feature library \mathbf{Q} , is zero-padded (gray regions) to ensure dimensional compatibility. It is assembled via column-wise concatenation over n_c identified material segments. The parameter vector $\boldsymbol{\theta} \in \mathbb{R}^{n_f n_c \times 1}$ encodes the model coefficients for each material segment. The right-hand side vector \mathbf{b} has two components: $\mathbf{b}^{\text{free}} = \mathbf{0}$ representing equilibrium at internal nodes, and \mathbf{b}^{fix} containing the boundary-aggregated reaction force.

$$\boldsymbol{\theta}^{\text{ols}} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{b} \quad (21)$$

However, such a solution would suffer from two main drawbacks – (i) the solution would be susceptible to overfitting and displacement noise [38] (see supplementary subsection S1.2), (ii) we would be unable to ensure that each component of $\boldsymbol{\theta}$ would be non-negative, which is essential to ensure polyconvexity of the strain energy density [46]. Thus, following [46] and [54], we formulate the equilibrium problem as a sparse-prior Bayesian regression for $\boldsymbol{\theta}$. The spike-slab prior [54] is used for $\boldsymbol{\theta}$ to enforce parsimony, avoid overfitting, and provide noise-robustness.

The equilibrium Equation 20 can be rewritten as

$$\mathbf{b} = \mathbf{A}\boldsymbol{\theta} + \epsilon$$

where $\epsilon \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$ is sampled from an independent and identical normal distribution with mean as $\mathbf{0}$ and variance as σ^2 . The above equation can be rewritten as

$$\mathbf{b} | \boldsymbol{\theta}, \sigma^2, \mathbf{A} \sim \mathcal{N}(\mathbf{A}\boldsymbol{\theta}, \sigma^2 \mathbf{I}) \quad (22)$$

Equation 22 forms a likelihood probability, and using Bayes' rule we can extract the posterior distribution $p(\boldsymbol{\theta}, \mathbf{z}, p_0, \nu_s, \sigma^2 | \mathbf{A}, \mathbf{b})$ using the approach outlined in Joshi et al. [46]. For brevity, we do not reiterate the Bayesian-EUCLID approach in the main text. A synopsis of the approach is provided in Appendix A. The solution for the material models $\boldsymbol{\theta}_{(k)} (k \in \{1 \dots N_G\})$ would be a collection of N_G number of values corresponding to the retained elements of the Markov chain, which approximate the posterior distribution for $\boldsymbol{\theta}$. For practical purposes, such as use in forward finite element simulations, it would be convenient to use the approximated mean of the distribution $\boldsymbol{\theta}_{\text{mean}} = \frac{\sum_{k \in \{1 \dots N_G\}} \boldsymbol{\theta}_{(k)}}{N_G}$ as material model parameters.

The standard deviation $\boldsymbol{\theta}_{\text{std}} = \sqrt{\frac{\sum_{k \in \{1 \dots N_G\}} (\boldsymbol{\theta}_{(k)} - \boldsymbol{\theta}_{\text{mean}})^2}{N_G}}$ provides an approximate measure of the

uncertainty in the prediction. The distribution for θ is multivariate, and θ_{std} provides only a practical, but approximate measure of the uncertainty associated with the material models. In this study, we provide the results as $\theta_{\text{mean}} \pm \theta_{\text{std}}$, instead of the violin plots in Joshi et al. [46].

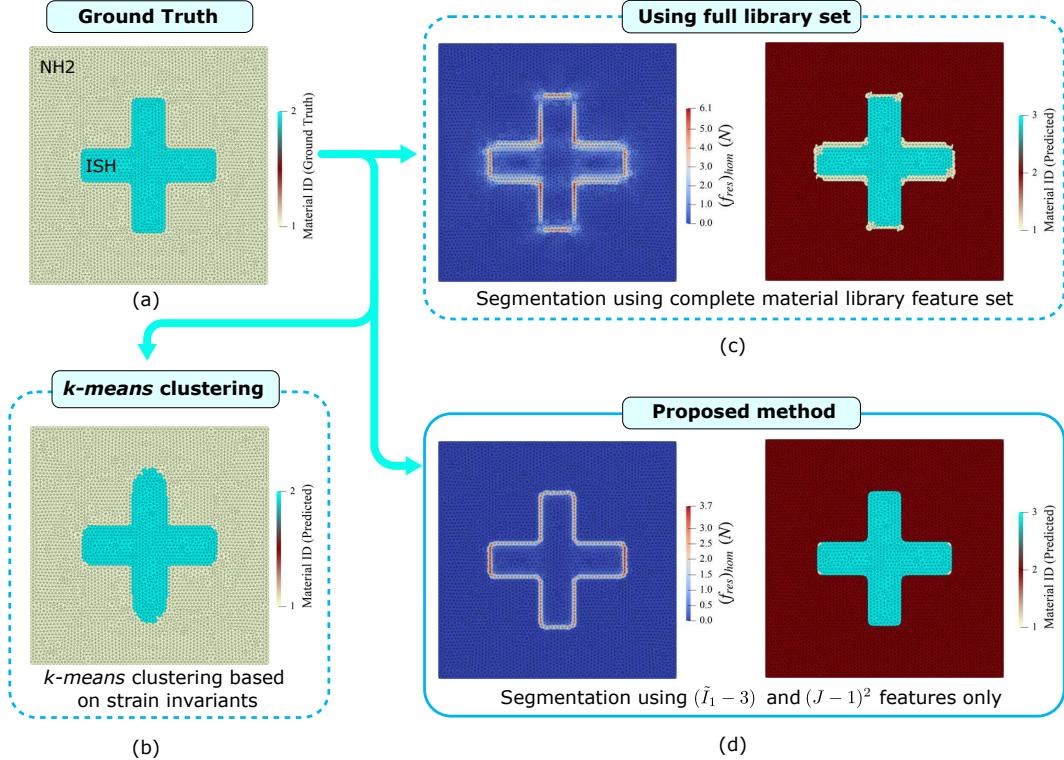


Figure 4: Comparative study of different segmentation techniques. (a) shows the ground truth for a heterogeneous bi-material specimen consisting of a cross-shaped inclusion governed by the Isihara (ISH) constitutive model, embedded in a base matrix described by the Neo-Hookean (NH2) model. (b) presents the segmentation result using k -means clustering based on strain invariants [55]. (c) utilizes the complete material feature library (Equation 6) to compute the 2D residual force norm distribution (f_{res}^a), and segments the material region by flagging nodes with residual force values above a specified threshold. (d) illustrates the result from our proposed method, where segmentation is performed using a reduced Neo-Hookean feature set ($\tilde{I}_1 - 3$) and $(J - 1)^2$, yielding clean and accurate material cluster identification in strong agreement with the ground truth.

2.2. Segmenting based on displacement data

Nguyen and Lejeune [55] provide a study on segmenting mechanically heterogeneous hyperelastic materials using surface displacement data. The isotropic invariants of the Cauchy-Green tensor (I_1, I_2 and I_3), alongside the displacement vector at the centroid of each finite element, were used as features to cluster each element using either k -means clustering or spectral clustering. This approach was validated on various heterogeneity patterns- circular inclusions, cross-shaped inclusions, Cahn-Hilliard patterns [56], and split domains. In most cases, their results found that the segmented regions approximated the underlying true segments and contained non-smooth boundaries. Furthermore, their results did not account for any correlated or uncorrelated noise in the displacement field, which could be expected to deteriorate the segmentation further. Subpart (b) of Figure 4 provides results of k -means clustering of strain invariants and displacements from a non-equibiaxial tension test on a square geometry with a cross-shaped inclusion. In Figure 4b, it can be seen that a significant number of elements are mis-clustered. Appendix B demonstrates that even 25 mis-clustered elements can deteriorate material property prediction accuracy by around 30%. Thus, any adopted segmentation approach should provide accurate boundaries in order to achieve reasonably accurate material property prediction.

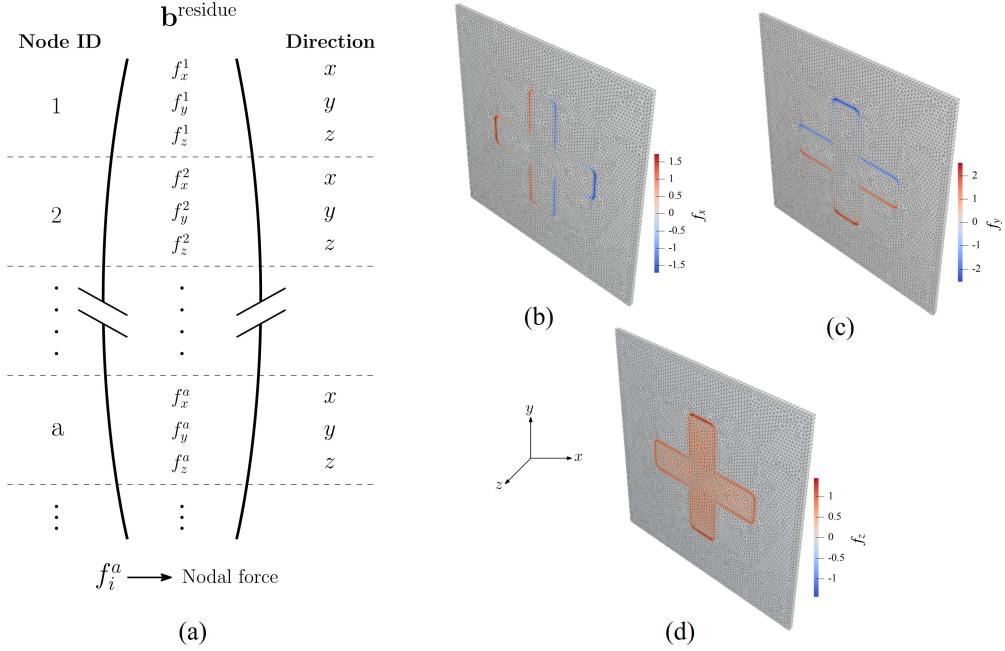


Figure 5: (a) Schematic of extracting nodal forces \mathbf{f}^a from $\mathbf{b}^{\text{residue}}$, (b)-(d) shows the three components of residual force.

In this work, we adopt two steps to achieve segmentation – the first step is to identify the inter-segment boundaries in the unknown material, and the next step is to use the identified boundaries to register all elements with their respective segment. The first step – identification of the inter-segment boundaries, is based on using “residual forces” within the geometry for a homogeneous model assumption. The \mathbf{A} (Equation 20) matrix is initially constructed assuming that the geometry contains only one segment (homogeneous) and only the NH2 features- ($I_1 - 3$) and ($J - 1$)² in the library. This is done only for segmenting the heterogeneous material to get a sharper inter-material interface (see Figure 4d), and in the subsequent Bayesian-EUCLID solution step, we keep all features in the library active. The homogeneous matrix \mathbf{A} constructed here is denoted by $\mathbf{A}_0 (\in \mathbb{R}^{(|\mathcal{D}^{\text{free}}|+3n_b) \times 2})$ to disambiguate it from the fully segmented \mathbf{A} matrix that will be used in the Bayesian-EUCLID step. As seen in the discussion preceding Equation 15, \mathbf{b}^{free} indicates the net force acting on each node within the volume of the material. For an appropriate guess of non-trivial material parameters ($\boldsymbol{\theta}^*$), we would have $\mathbf{A}^{\text{free}}\boldsymbol{\theta}^* = \mathbf{b}^{\text{free}} = \mathbf{0}$. However, no non-trivial solution exists for $\mathbf{A}^{\text{free}}\boldsymbol{\theta} = \mathbf{0}$. The Ordinary-Least Square (OLS) approximation $\boldsymbol{\theta}^{\text{ols}}$ (Equation 21) leaves the residue $\mathbf{A}^{\text{free}}\boldsymbol{\theta}^{\text{ols}} = \mathbf{b}^{\text{residue}} \neq \mathbf{b}^{\text{free}}$. Thus, we compute $\mathbf{b}^{\text{residue}}$ for the homogeneous material system \mathbf{A}_0 as follows

$$\mathbf{b}^{\text{residue}} = \mathbf{A}_0^{\text{free}} \left((\mathbf{A}_0^T \mathbf{A}_0)^{-1} \mathbf{A}_0^T \mathbf{b} \right) \quad (23)$$

For a node $a : (a, i) \in \mathcal{D}^{\text{free}}$, the nodal force (\mathbf{f}^a) along each direction is obtained using $\mathbf{b}^{\text{residue}}$ as shown in Figure 5. Thus, based on the homogenized model, we compute $\mathbf{b}^{\text{residue}}$ and then extract the residual nodal forces \mathbf{f}^a (see Figure 5a). We then make use of the residual nodal forces to demarcate the boundaries between two material segments by computing the 2D magnitude (ignoring f_z^a) as f_{res}

$$f_{\text{res}}^a = \sqrt{(f_x^a)^2 + (f_y^a)^2} \quad \forall \quad a \in \mathcal{N}_{\text{free}} \quad (24)$$

where,

$$\mathcal{N}_{\text{free}} = \{a : (a, i) \in \mathcal{D}^{\text{free}} \quad \forall \quad i \in \{1, 2, 3\}\} \quad (25)$$

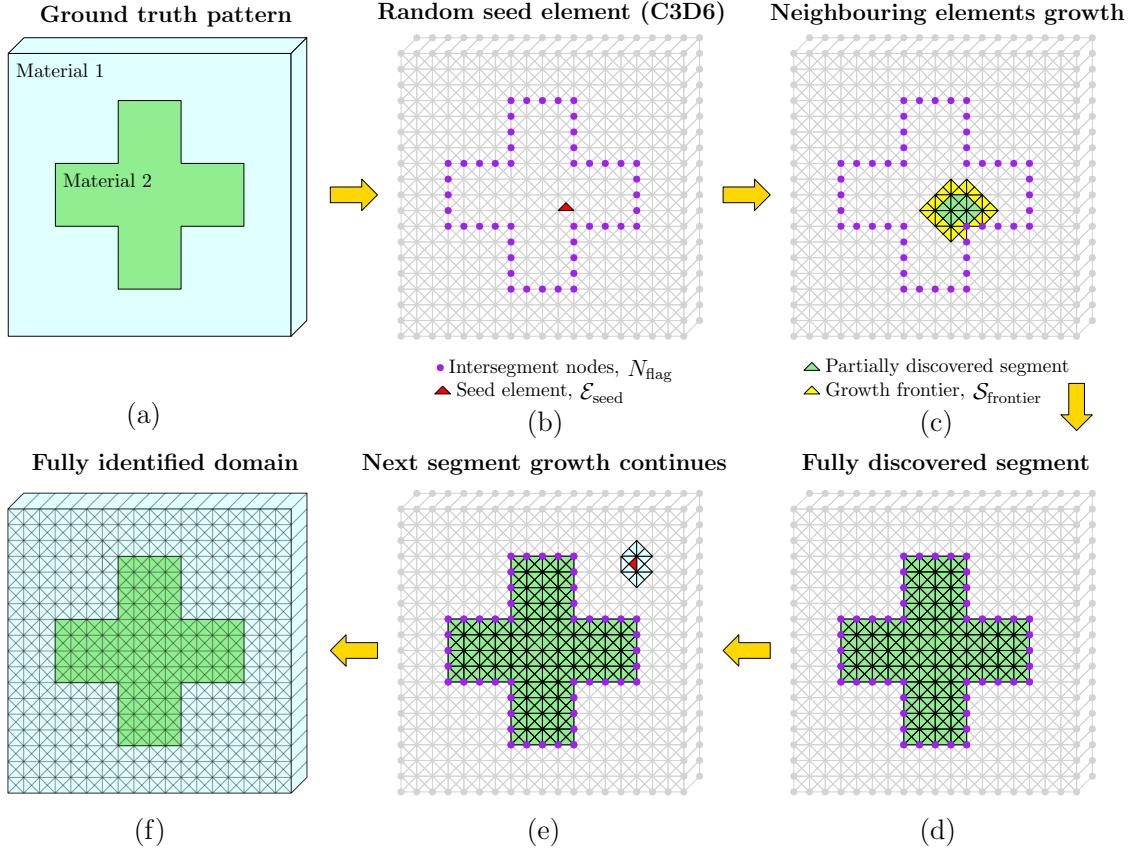


Figure 6: (a)-(f) Schematic illustration of the random seed-based growth algorithm for segmenting material sub-domains. The algorithm initiates from a randomly selected seed element (wedge type) and iteratively grows by adding neighboring elements until it reaches flagged interface boundaries. A new seed element is then chosen from the remaining unvisited elements, and the growth process is repeated until all distinct regions are identified, terminating at either interface or external boundaries.

Since we consider heterogeneity in a thin plate geometry, the material properties vary along the x and y directions (2D). The mismatch in material properties across inter-segment boundaries induces higher x and y components of residual forces on the inter-segment boundary nodes compared to nodes elsewhere in the domain. In contrast, the z components of the residual force result from material misidentification and are high throughout each segment. Therefore, it was observed that the use of 2D magnitude of the nodal forces provides a clearer inter-segment boundary as opposed to the 3D magnitude for the thin plate geometry considered in this study (See Figure 5b-d).

Nodes with the 2D force magnitude f_{res}^a , higher than a certain threshold are “flagged” as inter-segment boundary nodes. Generally, the magnitude of residual force is commensurate with the extent of mismatch between the “guessed” material property (homogeneous assumption) and the true underlying material property. At the nodes that lie on the interface of two (or more) material segments, a homogeneous material assumption would induce significantly more residual force due to the imbalance between elements with different material properties. Thus, these interface nodes would be characterized by a significantly higher than average (across all nodes) residual force. Therefore, we employ the following scheme to “flag” inter-segment nodes.

$$a \in \mathbf{N}_{\text{flag}}, \iff f_{\text{res}}^a > \lambda \sigma_{f_{\text{res}}} \quad (26)$$

where a is an integer, \mathbf{N}_{flag} denotes the set of all flagged nodes, λ is a tunable scalar (usually ranging between 1.5 and 2.5) and the standard deviation $\sigma_{f_{\text{res}}}$ is given by

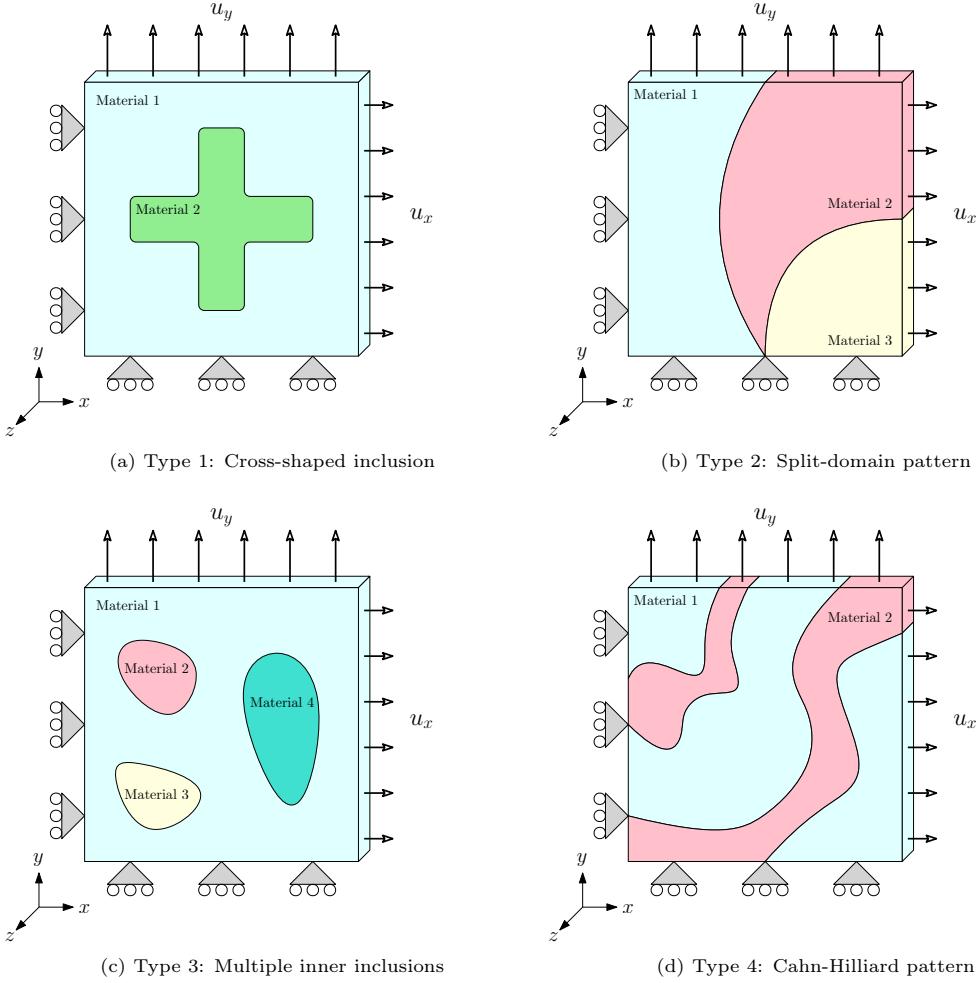


Figure 7: Graphical illustration of heterogeneous test specimen geometries subjected to non-equi-biaxial tensile loading under three-dimensional plane stress conditions. The heterogeneity patterns include: (a) cross-shaped inclusion, (b) split-domain configuration, (c) multiple non-circular inclusions, and (d) Cahn-Hilliard pattern.

$$\sigma_{f_{\text{res}}} = \sqrt{\frac{\sum_{a \in N_{\text{flag}}} (f_{\text{res}}^a - \mu)^2}{\sum_{a \in N_{\text{flag}}} 1}} \quad ; \quad \mu_{f_{\text{res}}} = \frac{\sum_{a \in N_{\text{flag}}} (f_{\text{res}}^a)}{\sum_{a \in N_{\text{flag}}} 1}$$

For zero and moderate noise, it was observed that the average of f_{res}^a , i.e., $\mu_{f_{\text{res}}}$ was significantly smaller than its standard deviation $\sigma_{f_{\text{res}}}$. However, $\mu_{f_{\text{res}}}$ was observed to increase with an increase in displacement noise, while $\sigma_{f_{\text{res}}}$ was nearly constant. This indicates that $\mu_{f_{\text{res}}}$ is a measure of the noise prevalent in the displacement measurement. Further, this implies that f_{res}^a could be high in nodes that are either on the inter-segment boundary or have high displacement noise. In order to disambiguate between the inter-segment nodes and those with high displacement noise, we employ a scalar multiple of $\sigma_{f_{\text{res}}}$ as a lower threshold. In experiments where the displacement noise is sufficiently low (measured by the ratio: $\mu_{f_{\text{res}}}/\sigma_{f_{\text{res}}}$), the inter-segment nodes can be expected to mostly have f_{res}^a above the lower threshold, while the "noisy" nodes would mostly have f_{res}^a below the threshold. This distinction blurs with increasing displacement noise, making residual force-based segmentation inaccurate for data with high displacement noise. In subsection 3.4 we provide further discussion on the observed statistical trends for nodal residual force f_{res}^a . Notably, Shi et al.

[45] makes similar use of residual forces from a homogenized assumption of their NN surrogate model to highlight the inter-subdomain boundaries, with user-defined lower thresholds.

Algorithm 1 Random seed-based linear element growth for material segmentation

Inputs: Flagged nodes \mathbf{N}_{flag} , Nodal positions, Connectivity
Definitions: $N_a \rightarrow$ Node a , with $a \in \{1 \dots n_{\text{nodes}}\}$
Definitions: $\mathcal{E}_i \rightarrow$ Element with ID i , with $i \in \{1 \dots n_{\text{els}}\}$ ▷ An element is a set of nodes
Definitions: $\mathcal{S}_i \rightarrow$ Material segment with ID i ▷ A segment is a set of elements
Definitions: $\mathcal{S}_{\text{domain}} \rightarrow$ Set of all elements in domain

```

1: function getMaterialSegments(Inputs)
2:    $\mathbf{N}_{\text{flag}} = \{N_a : (a, i) \in \mathcal{D}^{\text{free}} \quad \forall \quad i \in \{1, 2, 3\} \quad \& \quad f_{\text{res}}^a > \lambda \sigma_{f_{\text{res}}}^a\}$  ▷ Refer Equation 26 for  $\lambda$  and  $\sigma_{f_{\text{res}}}$ 
3:   Initialize  $i \leftarrow 1$ 
4:   Initialize  $\mathcal{S}_i \leftarrow \emptyset$ 
5:   while  $(\cup_{k=1}^i \mathcal{S}_k) \neq \mathcal{S}_{\text{domain}}$  do
6:     Select random seed element:  $\mathcal{E}_{\text{seed}} \in \mathcal{S}_{\text{domain}} \setminus (\cup_{k=1}^{i-1} \mathcal{S}_k)$ 
    With additional requirement:  $(\mathcal{E}_{\text{seed}}) \cap (N_{\text{flag}}) = \emptyset$ 
7:     if  $\mathcal{E}_{\text{seed}} == \emptyset$  then
8:       break loop
9:     end if
10:    Initialize growth frontier:  $\mathcal{S}_{\text{frontier}} \leftarrow \{\mathcal{E}_{\text{seed}}\}$ 
11:     $\mathcal{S}_i \leftarrow \mathcal{S}_{\text{frontier}}$ 
12:    while  $\mathcal{S}_{\text{frontier}} \neq \emptyset$  do
13:       $\mathcal{S}_{\text{neighbors}} \leftarrow \emptyset$ 
14:      for  $\mathcal{E}_k$  in  $\mathcal{S}_{\text{frontier}}$  do
15:        if  $\mathcal{E}_k \cap (N_{\text{flag}}) = \emptyset$  then
16:           $\mathcal{S}_{\text{neighbors}} \leftarrow \mathcal{S}_{\text{neighbors}} \cup \text{Neighbors}(\mathcal{E}_k)$ 
        ▷ Neighbors( $\mathcal{E}$ ) denotes the set of elements having at least one node in common with element  $\mathcal{E}$ 
17:        end if
18:      end for
19:       $\mathcal{S}_{\text{neighbors}} \leftarrow \mathcal{S}_{\text{neighbors}} \setminus (\mathcal{S}_{\text{neighbors}} \cap \mathcal{S}_i)$ 
20:       $\mathcal{S}_i \leftarrow \mathcal{S}_i \cup \mathcal{S}_{\text{neighbors}}$ 
21:       $\mathcal{S}_{\text{frontier}} \leftarrow \mathcal{S}_{\text{neighbors}}$ 
22:    end while
23:     $i \leftarrow i + 1$ 
24:    Initialize  $\mathcal{S}_i \leftarrow \emptyset$ 
25:  end while
   $i - 1$  indicates the number of discovered material segments
   $\mathcal{S}_k$  is the set of elements contained by the  $k^{\text{th}}$  material segment
Output:  $\{\mathcal{S}_k : k \in 1 \dots (i - 1)\}$ 
26: end function

```

After identifying the inter-segment boundaries, we employ an approach adapted and modified from Shi et al. [45] to register each element in the domain to a material segment. The approach is essentially a single-thread (cannot be parallelized) recursive growth algorithm. It involves choosing a seed element $\mathcal{E}_{\text{seed}}$ from anywhere within the domain ($\mathcal{S}_{\text{domain}}$), with a condition that none of the nodes constituting the element are “flagged” as inter-segment nodes (\mathbf{N}_{flag}). Next, the seed element and all its neighboring elements ($\mathcal{S}_{\text{neighbors}}$) are added into the material segment being discovered \mathcal{S}_i , with i denoting the index of the material segment. Next, those elements in the neighboring elements set $\mathcal{S}_{\text{neighbors}}$ that were previously absent in \mathcal{S}_i and do not contain flagged inter-segment nodes are designated as “growth-frontiers” ($\mathcal{S}_{\text{frontier}}$). The neighbors of each element belonging to $\mathcal{S}_{\text{frontier}}$ are then added to the material segment set \mathcal{S}_i , and some of these neighboring elements are then designated as growth frontiers and the growth continues till a

single material segment demarcated by a closed loop of flagged inter-segment nodes is completely discovered. The process then continues with the selection of another seed element from the set $\mathcal{S}_{\text{domain}} \setminus \mathcal{S}_i$ until all elements are registered to a material segment set \mathcal{S}_i . A more formal description of the approach is provided in Algorithm 1, and is depicted in Figure 6.

The important outcome of segmentation is the \mathbf{A} matrix as shown in Figure 3, with all elements and nodes mapped to appropriate material segments.

2.3. Robustness to noise

To evaluate the robustness of the Hetero-EUCLID framework, we introduce artificial noise into the synthetic surface displacement data obtained from finite element simulations. This was done to mimic experimental displacement measurement noise that would arise due to digital image correlation (DIC) and camera resolution limitations. For the specimen shown in Figure 7, the reference configuration measures $50 \text{ mm} \times 50 \text{ mm} \times 1 \text{ mm}$ and is subjected to non-equi-biaxial mechanical loading, with final stretch ratios of $\lambda_x = 1.6$ and $\lambda_y = 2.2$. The planar area of the final deformed plate would be $80 \times 110 \text{ mm}$. A typical DIC setup with a 1-megapixel camera would result in an approximate normalized field of view of 0.1 mm/pixel . As reported in Bornert et al. [57], the noise level in DIC setups usually varies between 0.01 to 0.05 pixels, which would correspond to $1 \times 10^{-3} \text{ mm}$ to $5 \times 10^{-3} \text{ mm}$ of noise for the geometry used in the current study. For the current study, we consider two cases of moderate and high noise level by introducing Gaussian noise having zero mean and standard deviation of $\sigma_{u,\text{mod}} = 2 \times 10^{-3} \text{ mm}$ and $\sigma_{u,\text{high}} = 5 \times 10^{-3} \text{ mm}$, respectively. Hence, the displacement field is modified as follows

$$u_i^a = (u_i^a)_{\text{FEM}} + (u_i^a)_{\text{noise}} \quad \text{with} \quad (u_i^a)_{\text{noise}} \sim \mathcal{N}(0, \sigma_u^2) \quad \forall \quad (a, i) \in \mathcal{D} \quad (27)$$

where $(u_i^a)_{\text{FEM}}$ denotes the measured input displacement field without noise and $(u_i^a)_{\text{noise}}$ represents the artificially introduced uncorrelated Gaussian noise at each node a and direction i . Next, as would be done with experimental data, the noisy displacement field \mathcal{U} is denoised using Kernel Ridge Regression (KRR) [58] having a Radial Basis Function (RBF) kernel with hyperparameters tuned by random search to minimize the standard deviation of the denoised signal.

Within the Bayesian-EUCLID framework for solving the constitutive model (represented by $\boldsymbol{\theta}$ in Equation 20), as was done in Joshi et al. [46], we implement a random sub-sampling of rows in \mathbf{A}^{free} , which is equivalent to sub-sampling nodes from the interior of the domain. The set of nodes \mathbf{N}_{free} and \mathbf{N}_{flag} were defined earlier in Equation 25 and in Algorithm 1, respectively. We subsample rows of \mathbf{A}^{free} corresponding to subsets of \mathbf{N}_{free} and \mathbf{N}_{flag} as follows

- **Free nodes:** For nodes in \mathbf{N}_{free} within the reference domain Ω , we randomly sample a subset of inner nodes such that $n_{\text{free}} < |\mathcal{D}^{\text{free}}|$. Typically, n_{free} is chosen as 2% of the total free degrees of freedom. However, in cases involving complex heterogeneity such as Cahn–Hilliard patterns with added displacement noise, the value of n_{free} may be increased up to 10% to ensure sufficient coverage.
- **Inter-segment nodes:** For nodes in \mathbf{N}_{flag} , we first solve for $\mathbf{b}_{\text{het}}^{\text{residue}}$ as follows

$$\mathbf{b}_{\text{het}}^{\text{residue}} = \mathbf{A}^{\text{free}} \left((\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{b} \right)$$

Next, we extract $(f_{\text{res}}^a)_{\text{het}}$ from $\mathbf{b}_{\text{het}}^{\text{residue}}$ as described in Equation 24 and Figure 5(a). We subsample $(f_{\text{res}}^a)_{\text{het}}$ for each inter-segment boundary and chose 20% of the nodes with the least $(f_{\text{res}}^a)_{\text{het}}$ for every inter-segment boundary. The use of this scheme for subsampling nodes from the inter-segment boundaries is expected to decrease the number of mislabeled (assigned wrong material segment) nodes contributing towards solving for $\boldsymbol{\theta}$, thereby improving its accuracy and noise-robustness. A detailed study regarding the effect of inter-segment node sub-sampling is provided in Appendix B.

The sub-sampling strategy described above significantly reduces the dimensionality of the matrices and vectors involved in Equation 20 thus significantly improving the computational speed. While the expression

in [Equation 19](#) remains unchanged, the system defined in [Equation 15](#) is modified to account for the reduced number of degrees of freedom as follows

$$\begin{aligned} \mathbf{A}^{\text{fix}}\boldsymbol{\theta} &= \mathbf{b}^{\text{fix}} \quad \text{with } \mathbf{A}^{\text{fix}} \in \mathbb{R}^{3n_b \times n_f n_c}, \quad \mathbf{b}^{\text{fix}} \in \mathbb{R}^{3n_b} \\ \mathbf{A}^{\text{free}}\boldsymbol{\theta} &= \mathbf{b}^{\text{free}} \quad \text{with } \mathbf{A}^{\text{free}} \in \mathbb{R}^{(n_{\text{free}}+n_{\text{flag}}) \times n_f n_c}, \quad \mathbf{b}^{\text{free}} \in \mathbb{R}^{(n_{\text{free}}+n_{\text{flag}})} \end{aligned} \quad (28)$$

[Equation 28](#) defines the updated force balance equations incorporating sub-sampled nodal contributions from free, flagged, and fixed regions of the domain.

3. Results and discussion

3.1. Computational data generation

To evaluate the proposed method, we perform finite element simulations using Abaqus/CAE 2025 software to generate a synthetic computational dataset corresponding to various heterogeneity patterns, as illustrated in [Figure 7](#). The model consists of a thin square plate measuring 50 mm × 50 mm × 1 mm, composed of multiple hyperelastic material segments and subjected to quasi-static, displacement-controlled non-equi-biaxial tension. The specimen is analyzed under the assumption of plane stress conditions. The domain is discretized using 6-node linear triangular prism elements (wedge or C3D6 type) as shown in [Figure 2](#). Simulations are conducted under prescribed boundary conditions, and both nodal displacements and boundary-aggregate reaction forces are recorded over a single load step. To emulate realistic DIC measurements, artificial Gaussian white noise is added to the displacement field, as described in [Equation 27](#), with subsequent noise handling procedures detailed in [subsection 2.3](#).

For each heterogeneity pattern shown in [Figure 7](#), we assign different combinations of hyperelastic material models that include a quadratic volumetric strain energy term. These configurations are used to benchmark the performance of the Hetero-EUCLID framework. For the scope of the current work, we use the material library \mathbf{Q} with six features as highlighted in [Table 1](#). We employ only $n_f = 6$ features in the library instead of 24 features in the previous Bayesian-EUCLID work [46]. This is because the number of columns of \mathbf{A} matrix would be $n_f n_c = 30$, if there are $n_c = 5$ material segments in the domain. Using a higher number of features (n_f) would reduce the speed of computation of the Markov Chain solution and is therefore not attempted in this work. The various hyperelastic material models considered in this work are

- A. Neo-Hookean model (NH2) [49]

$$W_{\text{NH2}}(\mathbf{C}; \boldsymbol{\theta}) = \theta_1(\tilde{I}_1 - 3) + \theta_6(J - 1)^2 \quad (29)$$

- B. Isihara model (ISH) [48]

$$W_{\text{ISH}}(\mathbf{C}; \boldsymbol{\theta}) = \theta_1(\tilde{I}_1 - 3) + \theta_2(\tilde{I}_2 - 3) + \theta_3(\tilde{I}_1 - 3)^2 + \theta_6(J - 1)^2 \quad (30)$$

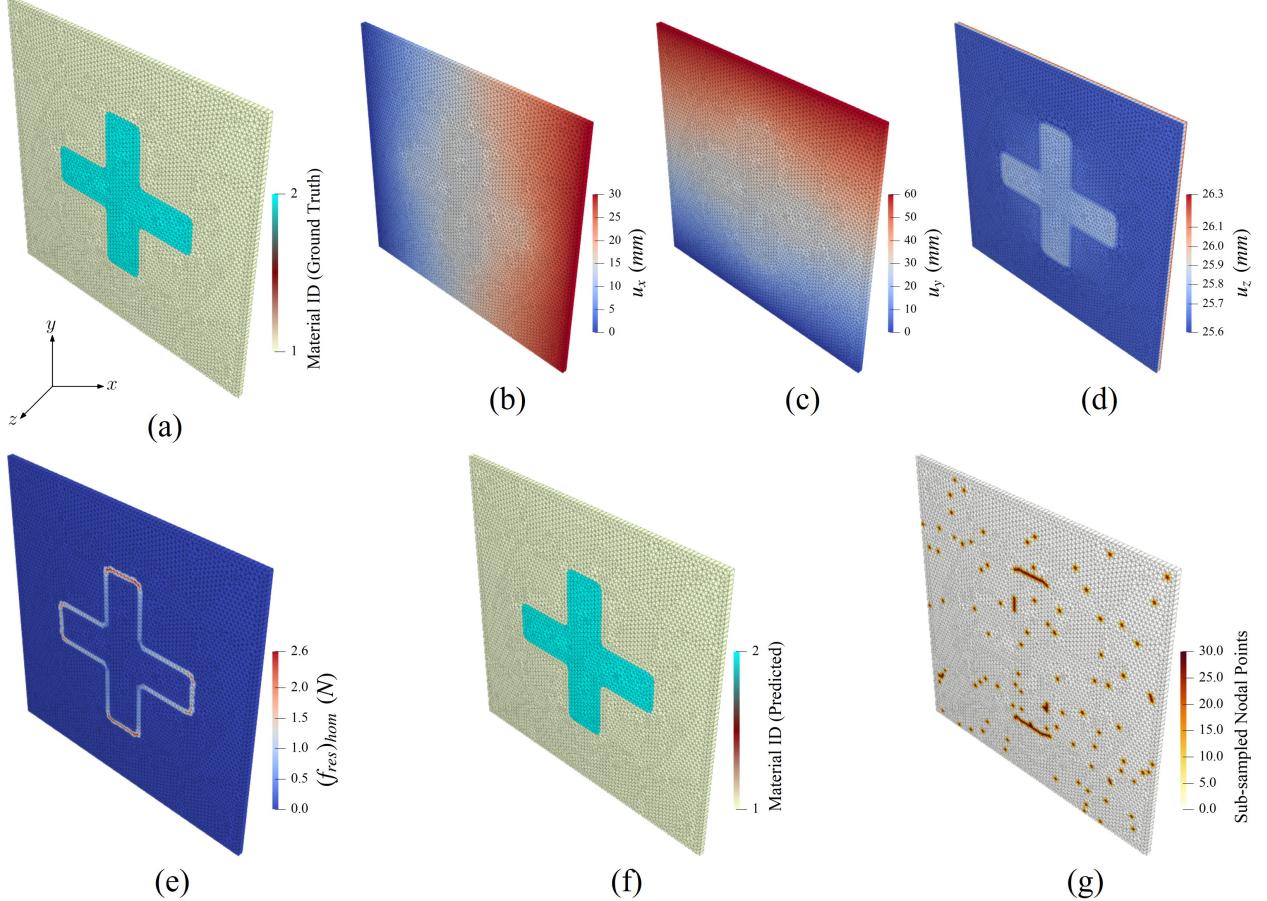
- C. Haines-Wilson model (HW) [51]

$$W_{\text{HW}}(\mathbf{C}; \boldsymbol{\theta}) = \theta_1(\tilde{I}_1 - 3) + \theta_2(\tilde{I}_2 - 3) + \theta_4(\tilde{I}_1 - 3)(\tilde{I}_2 - 3) + \theta_5(\tilde{I}_1 - 3)^3 + \theta_6(J - 1)^2 \quad (31)$$

Model coefficient	Library feature	Model coefficient	Library feature
θ_1	$(\tilde{I}_1 - 3)$	θ_4	$(\tilde{I}_1 - 3)(\tilde{I}_2 - 3)$
θ_2	$(\tilde{I}_2 - 3)$	θ_5	$(\tilde{I}_1 - 3)^3$
θ_3	$(\tilde{I}_1 - 3)^2$	θ_6	$(J - 1)^2$

Table 1: List of features in the library set \mathbf{Q} used to represent the Neo-Hookean (29), Isihara (30), and Haines-Wilson (31) constitutive models.

3.2. Pattern case study



Pattern Type: Cross-shaped inclusion | Mesh elements (C3D6): 10,442

Case	Material ID	Constitutive model	θ_1 (MPa)	θ_2 (MPa)	θ_3 (MPa)	θ_4 (MPa)	θ_5 (MPa)	θ_6 (MPa)	Free nodes sub-sampling
Ground Truth, (a)	1	NH2 _a	1.80	—	—	—	—	6.00	—
	2	NH2 _b	5.40	—	—	—	—	15.00	
No Noise, (f)	1	NH2 _a	1.79 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	6.00 ± 0.00	2%
	2	NH2 _b	5.46 ± 0.01	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	15.13 ± 0.04	

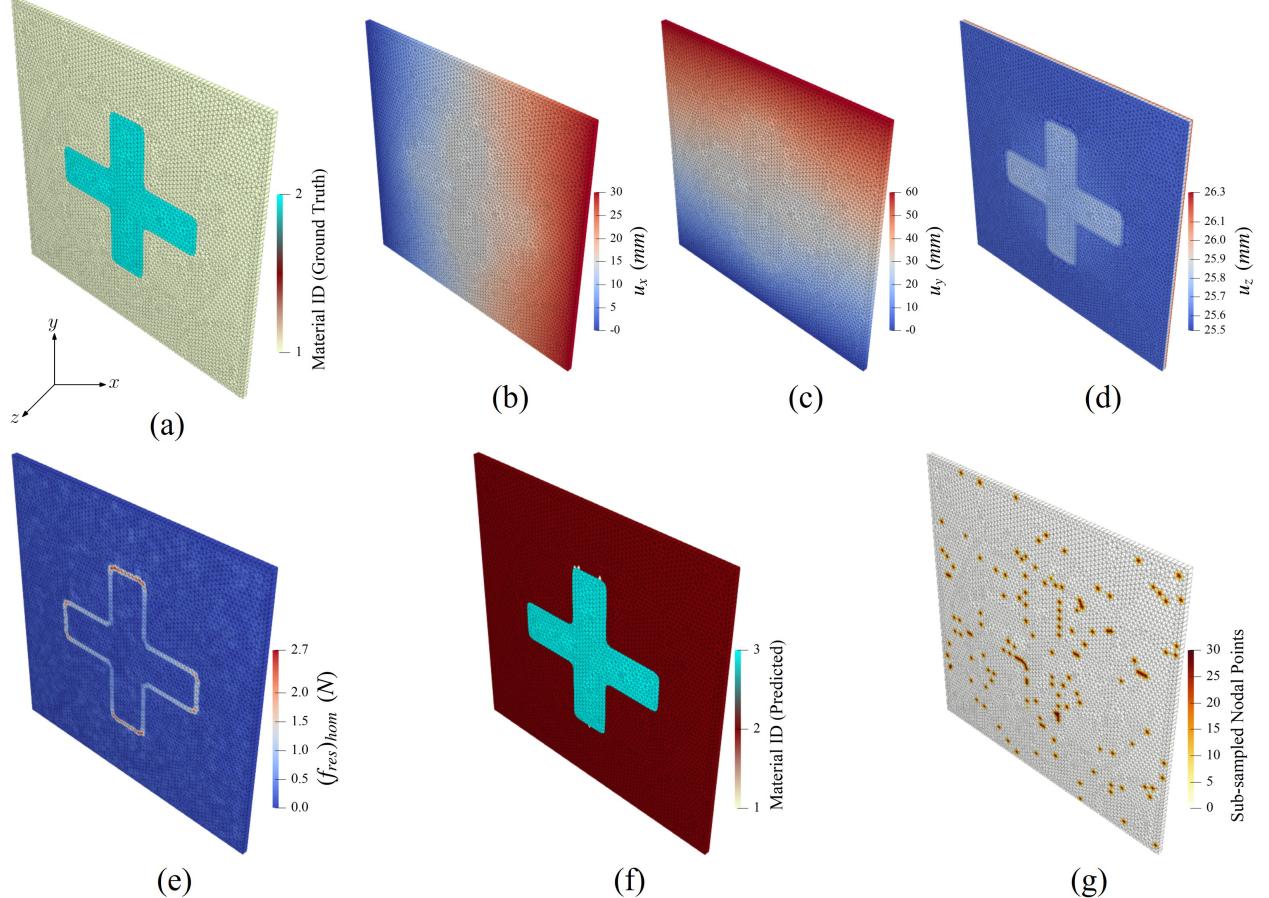
Table 2: True vs. predicted material parameters for cross-shaped inclusion pattern without noise using Bayesian-EUCLID framework.

Figure 8: Model discovery for the cross-shaped inclusion pattern (Figure 7a), with sub-figures (b)–(g) corresponding to the zero noise case. (a) displays the ground truth material segment IDs. (b)–(d) show the displacement fields in the x , y , and z directions, respectively. (e) presents the residual force norm distribution (Equation 24) from the homogenized model. (f) illustrates the predicted material segments from the segmentation algorithm. (g) highlights the sub-sampled nodes from the D^{free} subset used to construct the heterogenized model, which is subsequently solved using the Bayesian framework.

Type 1: Cross-shaped inclusion

Figure 7a illustrates a heterogeneous specimen with a cross-shaped inclusion embedded in a surrounding matrix material. Through this configuration we aim to examine the effectiveness of the Hetero-EUCLID framework; however, it presents a non-trivial test case due to the presence of sharp corners in inter-segment interfaces. For the cross-shaped inclusion pattern, we begin the analysis for a hyperelastic bi-material

setup where both the inclusion and the matrix follow Neo-Hookean constitutive models (NH2_a – NH2_b). The domain, modeled as a single-element-thick plate, is discretized using 10,442 uniformly sized wedge-type elements. As depicted in Figure 7a, the specimen is subjected to quasi-static, displacement-controlled non-equi-biaxial tensile loading over $n_t = 6$ time steps, reaching final directional stretch ratios of $\lambda_x = 1.6$ and $\lambda_y = 2.2$. Only displacement data from the final load-step $n_t = 6$ is considered for further analysis.



Pattern Type: Cross-shaped inclusion | Mesh elements (C3D6): 10,442

Case	Material ID	Constitutive model	θ_1 (MPa)	θ_2 (MPa)	θ_3 (MPa)	θ_4 (MPa)	θ_5 (MPa)	θ_6 (MPa)	Free nodes sub-sampling
Ground Truth, (a)	1	NH2_a	1.80	—	—	—	—	6.00	—
	2	NH2_b	5.40	—	—	—	—	15.00	
Moderate Noise $\sigma_u = 2 \times 10^{-3}$	1	NH2_a	1.81 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	6.01 ± 0.00	2%
	2	NH2_b	5.42 ± 0.07	0.00 ± 0.03	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	14.91 ± 0.14	
High Noise $\sigma_u = 5 \times 10^{-3}$, (f)	2	NH2_a	1.80 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	5.99 ± 0.00	2%
	3	NH2_b	5.25 ± 0.09	0.00 ± 0.01	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	14.40 ± 0.23	

Table 3: True vs. predicted material parameters for cross-shaped inclusion pattern with noise using Bayesian-EUCLID framework.

Figure 9: Model discovery for the cross-shaped inclusion pattern (Figure 7a), with sub-figures (b)–(g) corresponding to the high noise case. (a) displays the ground truth material segment IDs. (b)–(d) show the displacement fields in the x , y , and z directions, respectively. (e) presents the residual force norm distribution (Equation 24) from the homogenized model. (f) illustrates the predicted material segments from the segmentation algorithm. (g) highlights the sub-sampled nodes from the $\mathcal{D}^{\text{free}}$ subset used to construct the heterogenized model, which is subsequently solved using the Bayesian framework.

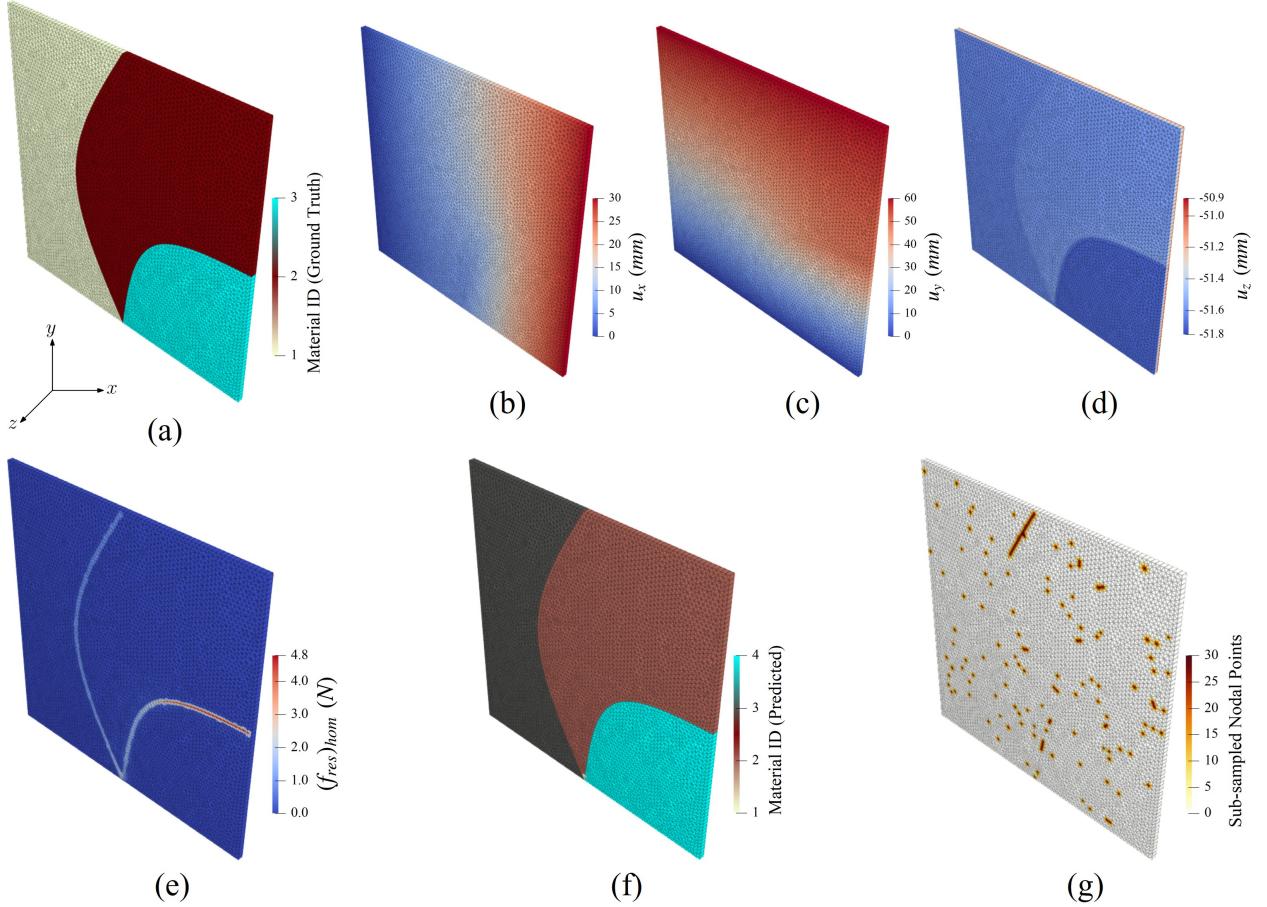
[Figure 8](#) summarizes the performance of Hetero-EUCLID in identifying interpretable material models for the cross-shaped inclusion specimen. [Figure 8a](#) shows the ground truth, indicating both the number of material segments and their true geometrical boundaries. The full-field displacement fields obtained from FEM simulations are visualized in [Figure 8b-d](#). Subsequently, solving the homogenized model using ordinary least squares regression yields the residual force distribution f_{res}^a , as shown in [Figure 8e](#). The presence of high residual force resembles the imbalance at the material interface and aids in flagging the interface nodes based on the thresholding procedure described in [Equation 26](#). Next, the domain is segmented using the random seed-based island growth algorithm described in [subsection 2.2](#). [Figure 8f](#) illustrates clean and precise segmentation of the heterogeneous pattern within the specimen. The algorithm systematically identifies both the number of distinct material sub-domains and their corresponding element sets. Using the segmentation results, the heterogenized model is assembled using [Equation 20](#) (refer to [Figure 3](#)), with nodal data selected via random sub-sampling as outlined in [subsection 2.3](#). The nodes included for solving [Equation 28](#) using the Bayesian-EUCLID framework are highlighted in [Figure 8g](#).

We then solve the system using a Bayesian framework with spike-slab priors, employing three parallel MCMC chains (Markov Chain Monte Carlo) of length 500 each (refer [Appendix A](#) and Joshi et al. [46] for details on the Bayesian framework). The outcome of this procedure is a segment-wise interpretable discovery of material models. Table 2 benchmarks the predicted model coefficients θ_1 and θ_6 against the ground truth values for the NH_{2a}-NH_{2b} material combination ([Equation 29](#)). It can be observed that, with only 2% of the domain nodal data sampled, the identified physically interpretable coefficients as shear and bulk moduli closely match their true values. For coefficients θ_2 to θ_5 , which are not expected to contribute under the Neo-Hookean model formulation, the Bayesian framework predicts zero values due to the sparsity-promoting nature of the spike-slab prior. The small standard deviation values further indicate high confidence in the inferred parameters.

The influence of noise on Hetero-EUCLID for cross-shaped inclusion is reported in [Figure 9](#). Under high noise case ($\sigma_{u,\text{high}} = 5 \times 10^{-3}$), the effect on the residual force norm (f_{res}^a) is shown in [Figure 9e](#), which may be visually compared with the noise-free case in [Figure 8e](#). Despite the added noise, the material region segmentation remains accurate as illustrated in [Figure 9f](#). Supplementary Figure S1 shows the corresponding visualization results for moderate noise condition ($\sigma_{u,\text{mod}} = 2 \times 10^{-3}$). Table 3 reports the predicted material parameters for both moderate and high noise cases, where the maximum absolute error remains below 1.2% compared to the ground truth. Supplementary section S2 presents an empirical validation study by comparing the strain energy density along six different deformation paths between the discovered material model and the corresponding ground truth. Supplementary Figure S5 shows the comparative strain energy density plots for the cross-shaped inclusion pattern under high noise conditions. The results indicate good agreement between the predicted and true strain energy densities, demonstrating the robustness of the proposed method under noise conditions.

Type 2: Split-domain pattern

We further investigate a more complex heterogeneous configuration, featuring a split-domain pattern with three distinct material regions whose interfaces converge at a common junction located on one of the external boundaries, as illustrated in [Figure 7b](#). The domain is discretized using 10,972 wedge elements, and the underlying ground truth material model comprises of a combination of Neo-Hookean, Isihara, and Haines-Wilson constitutive laws. Using the same non-equi-biaxial loading conditions as in the previous case, we perform inverse parameter identification with the Hetero-EUCLID framework under noise-free, high, and moderate noise conditions. The corresponding results are presented in [Figure 10](#), [11](#), and Supplementary Figure S2, respectively. In each of the above scenarios, the residual force distribution (f_{res}^a) reveals clear mechanical imbalance at the material interfaces, as shown in [Figure 10e](#), [11e](#) and S2e. Next, the segmentation step is executed using the seed-based growth algorithm with a tunable threshold parameter λ (refer to [Equation 26](#)), and the resulting predicted material segments are shown in [Figure 10f](#), [11f](#), and S2f. In each case, a small degree of misclustering is observed at the junction of the material interfaces, affecting only a few elements.

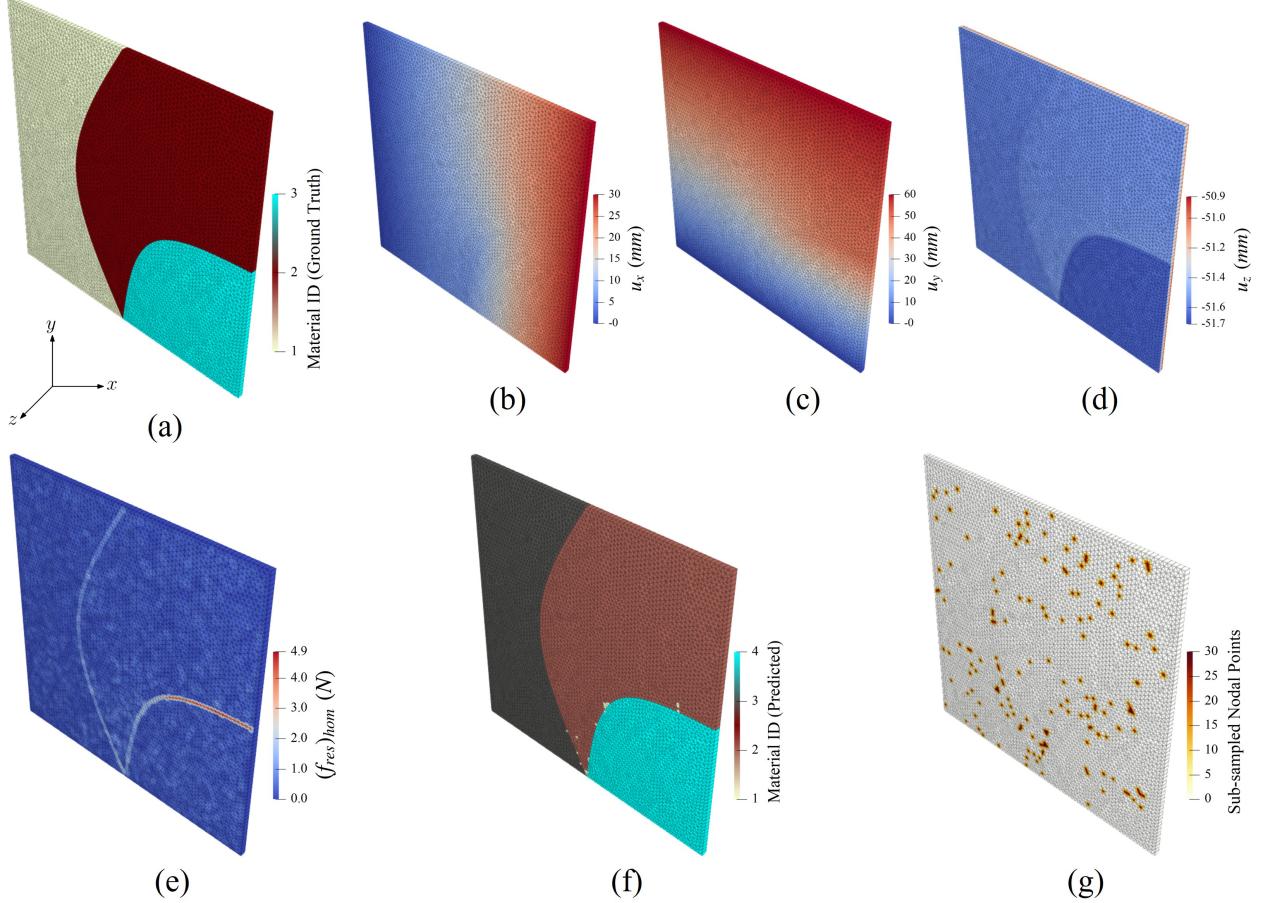


Pattern Type: Split-domain | Mesh elements (C3D6): 10,972

Case	Material ID	Constitutive model	θ_1 (MPa)	θ_2 (MPa)	θ_3 (MPa)	θ_4 (MPa)	θ_5 (MPa)	θ_6 (MPa)	Free nodes sub-sampling
Ground Truth Refer (a)	1	NH2	6.00	—	—	—	—	32.00	—
	2	ISH	4.00	0.50	0.30	—	—	21.00	—
	3	HW	1.00	0.15	—	0.02	0.00	10.00	—
No Noise Refer (f)	3	NH2	6.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	32.00 ± 0.00	—
	2	ISH	3.99 ± 0.03	0.51 ± 0.01	0.29 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	20.99 ± 0.01	2%
	4	HW	0.93 ± 0.09	0.21 ± 0.02	0.00 ± 0.01	0.02 ± 0.00	0.00 ± 0.00	10.05 ± 0.01	—

Table 4: True vs. predicted material parameters for split-domain pattern without noise using Bayesian-EUCLID framework.

Figure 10: Model discovery for the split-domain pattern (Figure 7b), with sub-figures (b)–(g) corresponding to the zero noise case. (a) displays the ground truth material segment IDs. (b)–(d) show the displacement fields in the x , y , and z directions, respectively. (e) presents the residual force norm distribution (Equation 24) from the homogenized model. (f) illustrates the predicted material segments from the segmentation algorithm. (g) highlights the sub-sampled nodes from the $\mathcal{D}^{\text{free}}$ subset used to construct the heterogenized model, which is subsequently solved using the Bayesian framework.



Pattern Type: Split-domain | Mesh elements (C3D6): 10,972

Case	Material ID	Constitutive model	θ_1 (MPa)	θ_2 (MPa)	θ_3 (MPa)	θ_4 (MPa)	θ_5 (MPa)	θ_6 (MPa)	Free nodes sub-sampling
Ground Truth Refer (a)	1	NH2	6.00	—	—	—	—	32.00	—
	2	ISH	4.00	0.50	0.30	—	—	21.00	
	3	HW	1.00	0.15	—	0.02	0.00	10.00	
Moderate Noise ($\sigma_u = 2 \times 10^{-3}$)	3	NH2	5.98 ± 0.02	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	31.79 ± 0.03	2%
	2	ISH	5.25 ± 0.16	0.20 ± 0.05	0.21 ± 0.02	0.00 ± 0.00	0.00 ± 0.00	21.89 ± 0.05	
	4	HW	1.62 ± 0.05	0.28 ± 0.01	0.00 ± 0.01	0.02 ± 0.00	0.00 ± 0.00	9.72 ± 0.03	
High Noise ($\sigma_u = 5 \times 10^{-3}$) Refer (f)	3	NH2	5.98 ± 0.03	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	31.62 ± 0.05	2%
	2	ISH	6.14 ± 0.20	0.00 ± 0.01	0.07 ± 0.08	0.00 ± 0.00	0.01 ± 0.01	21.80 ± 0.08	
	4	HW	1.70 ± 0.05	0.25 ± 0.01	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	9.57 ± 0.04	

Table 5: True vs. predicted material parameters for split-domain pattern with noise using Bayesian-EUCLID framework.

Figure 11: Model discovery for the split-domain pattern (Figure 7b), with sub-figures (b)–(g) corresponding to the high noise case. (a) displays the ground truth material segment IDs. (b)–(d) show the displacement fields in the x , y , and z directions, respectively. (e) presents the residual force norm distribution (Equation 24) from the homogenized model. (f) illustrates the predicted material segments from the segmentation algorithm. (g) highlights the sub-sampled nodes from the $\mathcal{D}^{\text{free}}$ subset used to construct the heterogenized model, which is subsequently solved using the Bayesian framework.

Table 4 compares the predicted material parameters $\boldsymbol{\theta}$ under noise-free conditions against the ground truth values. The parameters corresponding to the Neo-Hookean (NH2: $\theta_{1,6}$), Isihara (ISH: $\theta_{1-3,6}$), and

Haines–Wilson (HW: $\theta_{1,2,4–6}$) models are all recovered with high accuracy and low standard deviation. Under noisy conditions, as reported in Table 5, the region governed by the NH2 model is predicted accurately. However, the regions corresponding to ISH and HW models exhibit some deviations in the predicted parameters, likely due to the compounded effect of noise and minor misclustering. The noise-induced deviation can also be observed from the strain energy density plots for HW under high noise conditions, as shown in Supplementary Figure S6. The prediction accuracies can be expected to improve with the use of data from multiple experiments/load-steps instead of a single non-equi-biaxial stretch test.

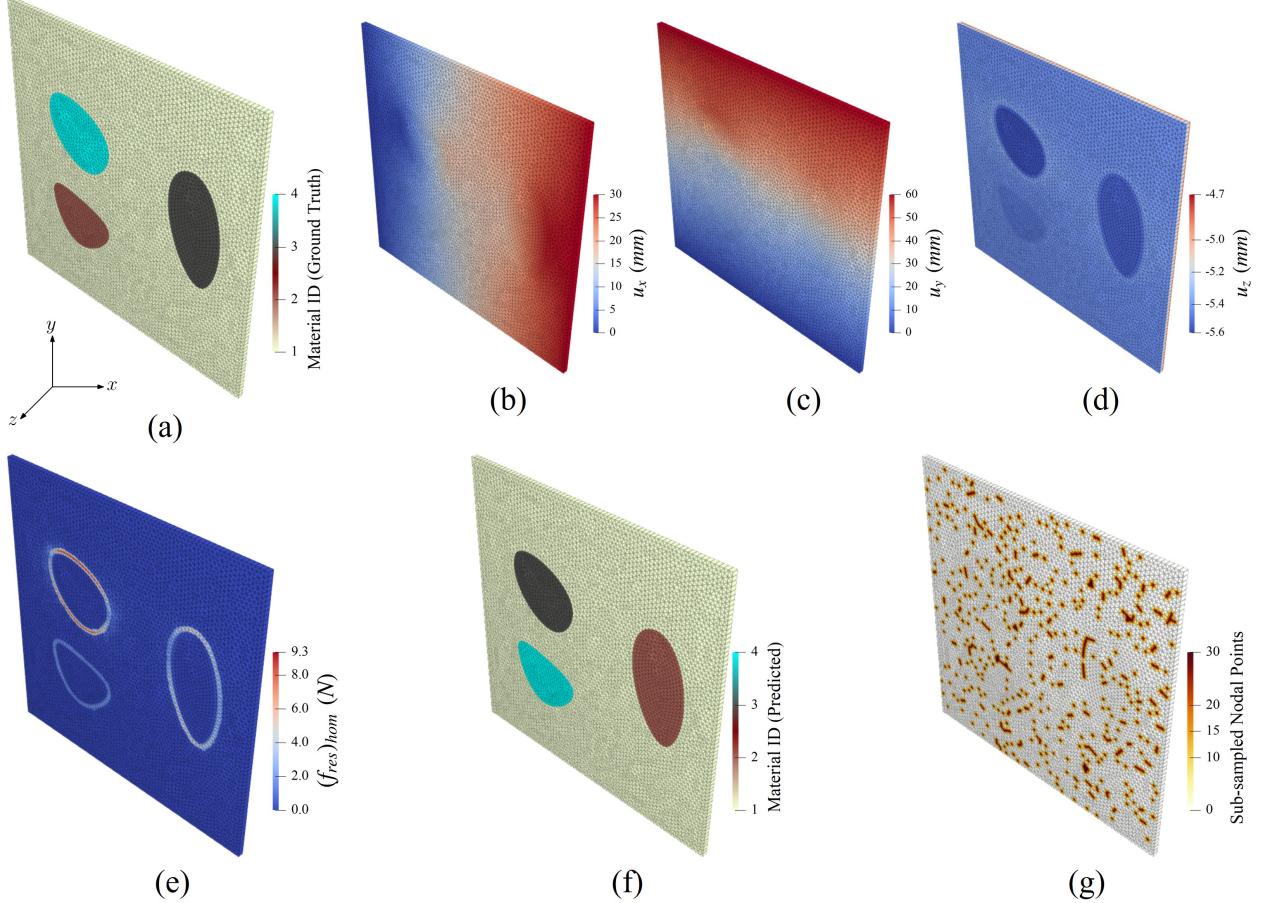
Type 3: Multiple inner inclusions

Next, we extend our investigation to a heterogeneous configuration featuring multiple inclusions with non-circular geometries. [Figure 7c](#) illustrates the case where the matrix follows an Isihara (ISH) material model, and three embedded inclusions are governed by Neo-Hookean (NH2_a, NH2_b) and Haines–Wilson (HW) constitutive laws. Using this type of configuration we aim to evaluate the proposed framework’s ability to simultaneously segment and characterize multiple non-circular inclusions with varying classes of material models. The domain is discretized using 10,816 uniformly spaced wedge-type elements and subjected to the same non-equi-biaxial displacement-controlled loading conditions used in previous cases. We investigated the Type 3 specimen for moderate and high noise levels. [Figure 12](#) presents the results of material region segmentation and parameter identification under noise-free conditions. Based on the residual force distribution shown in [Figure 12e](#), and applying the random seed-based island growth algorithm, the predicted material segments align well with the true inclusions, as illustrated in [Figure 12f](#). Table 6 reports the accurate recovery of material parameters for all regions in the ISH–NH2_a–HW–NH2_b configuration, with low standard deviations indicating high confidence in the prediction.

Under high-noise conditions ($\sigma_{u,\text{high}} = 5 \times 10^{-3}$), it can be observed that material region prediction of the embedded inclusions remains accurate with a few omissions of elements from the base matrix as illustrated in [Figure 13f](#). Supplementary Figure S3 displays the corresponding results with moderate noise condition ($\sigma_{u,\text{mod}} = 2 \times 10^{-3}$). Table 7 summarizes the identified material parameters for each region under the moderate and high-noise settings. For segments corresponding to ISH and NH2_a materials (material IDs 2 and 5), the predicted parameters $\theta_{1–6}$ remain close to their ground truth values. In contrast, the material parameter prediction for the HW and NH2_b regions (material segment IDs 3 and 4, respectively) show large deviation from the ground truth with higher standard deviations, indicating a possible breakdown of the proposed method under strong noise interference. Supplementary Figure S7 compares the predicted strain energy density along six distinct deformation paths with the ground truth for the ISH–NH2_a–HW–NH2_b configuration for high-noise. The plots show that the energy density predictions for the ISH and NH2_a segments exhibit higher accuracy, whereas the predictions for HW and NH2_b show lower R^2 scores and wider percentile uncertainty bands. The result suggests that additional experimental data, particularly under varied loading conditions, may be necessary to better capture the shear-dominated material response as the volumetric stiffnesses corresponding to θ_6 are accurately predicted for each material.

Type 4: Cahn–Hilliard pattern

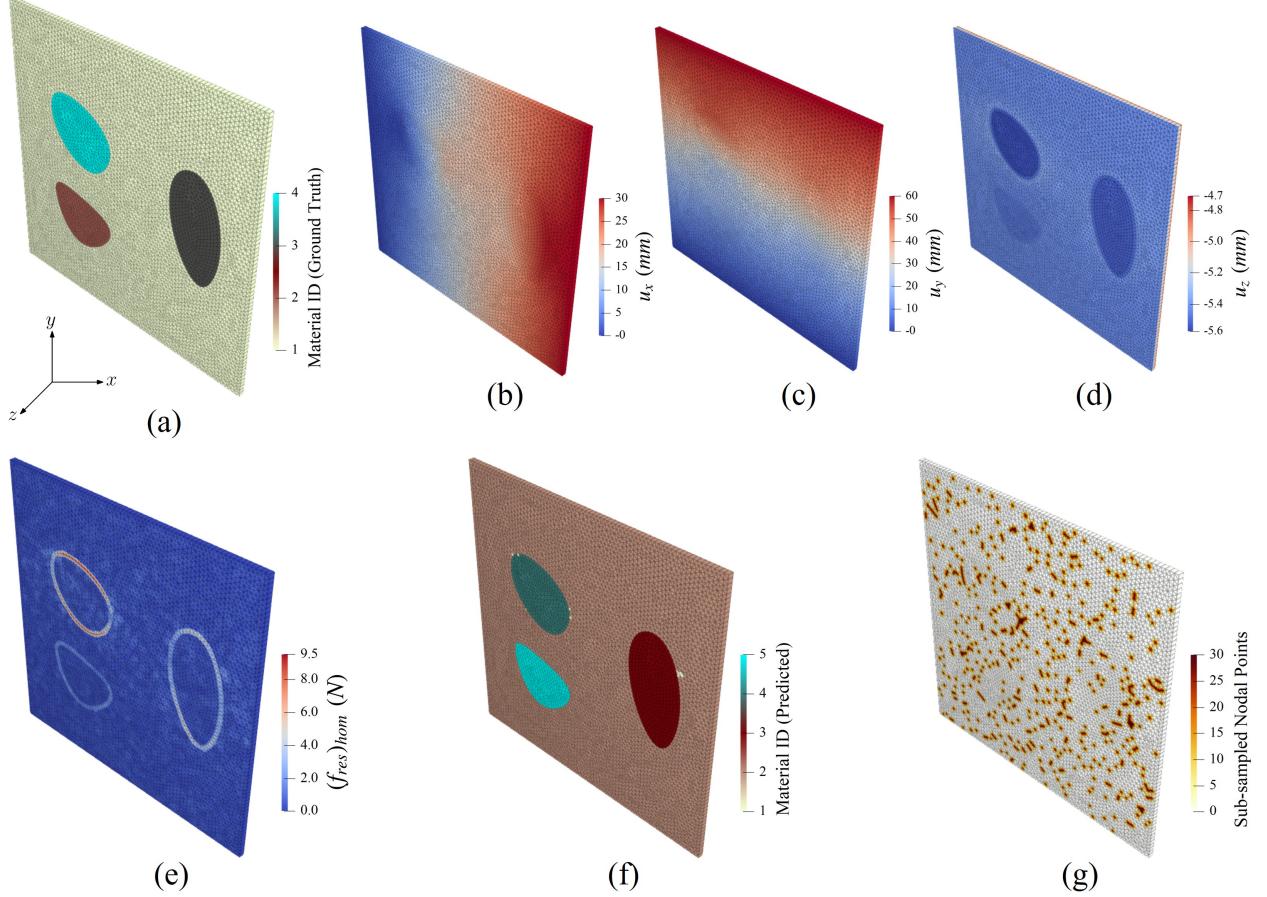
In this type of domain heterogeneity, we investigate the Cahn–Hilliard pattern, which is often used to represent the biological pattern formation in soft tissue mechanics. [Figure 7d](#) illustrates this configuration, featuring five adjacent material regions alternately composed of two Neo-Hookean (NH2_a, NH2_b) constitutive models. This specimen serves to evaluate the ability of Hetero-EUCLID to distinguish material sub-regions when multiple zones share identical mechanical properties. The domain is meshed with 21,040 uniformly distributed wedge-type elements. Using ABAQUS/CAE, we generate the full-field computational displacement dataset for quasi-static non-equi-biaxial loading as shown in [Figure 14b-d](#). The Cahn–Hilliard heterogeneous configuration is analyzed under noise-free, high, and moderate-noise conditions, with the corresponding results summarized in [Figure 14](#), [15](#), and Supplementary Figure S4, respectively.



Pattern Type: Multiple inner inclusions Mesh elements (C3D6): 10,816									
Case	Material ID	Constitutive model	θ_1 (MPa)	θ_2 (MPa)	θ_3 (MPa)	θ_4 (MPa)	θ_5 (MPa)	θ_6 (MPa)	Free nodes sub-sampling
Ground Truth Refer (a)	1	ISH	4.00	0.50	0.30	—	—	21.00	—
	2	NH2 _a	5.40	—	—	—	—	15.00	—
	3	HW	1.00	0.15	—	0.02	0.00	10.00	—
	4	NH2 _b	1.80	—	—	—	—	6.00	—
No Noise Refer (f)	1	ISH	4.00 ± 0.00	0.50 ± 0.00	0.30 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	21.01 ± 0.00	—
	4	NH2 _a	5.40 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	14.99 ± 0.01	10%
	2	HW	0.72 ± 0.07	0.00 ± 0.03	0.00 ± 0.03	0.00 ± 0.00	0.00 ± 0.00	9.81 ± 0.01	—
	3	NH2 _b	1.75 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	5.83 ± 0.01	—

Table 6: True vs. predicted material parameters for multiple inner inclusions pattern without noise using Bayesian-EUCLID framework.

Figure 12: Model discovery for the multiple inner inclusions pattern (Figure 7c), with sub-figures (b)–(g) corresponding to the zero noise case. (a) displays the ground truth material segment IDs. (b)–(d) show the displacement fields in the x , y , and z directions, respectively. (e) presents the residual force norm distribution (Equation 24) from the homogenized model. (f) illustrates the predicted material segments from the segmentation algorithm. (g) highlights the sub-sampled nodes from the D^{free} subset used to construct the heterogenized model, which is subsequently solved using the Bayesian framework.

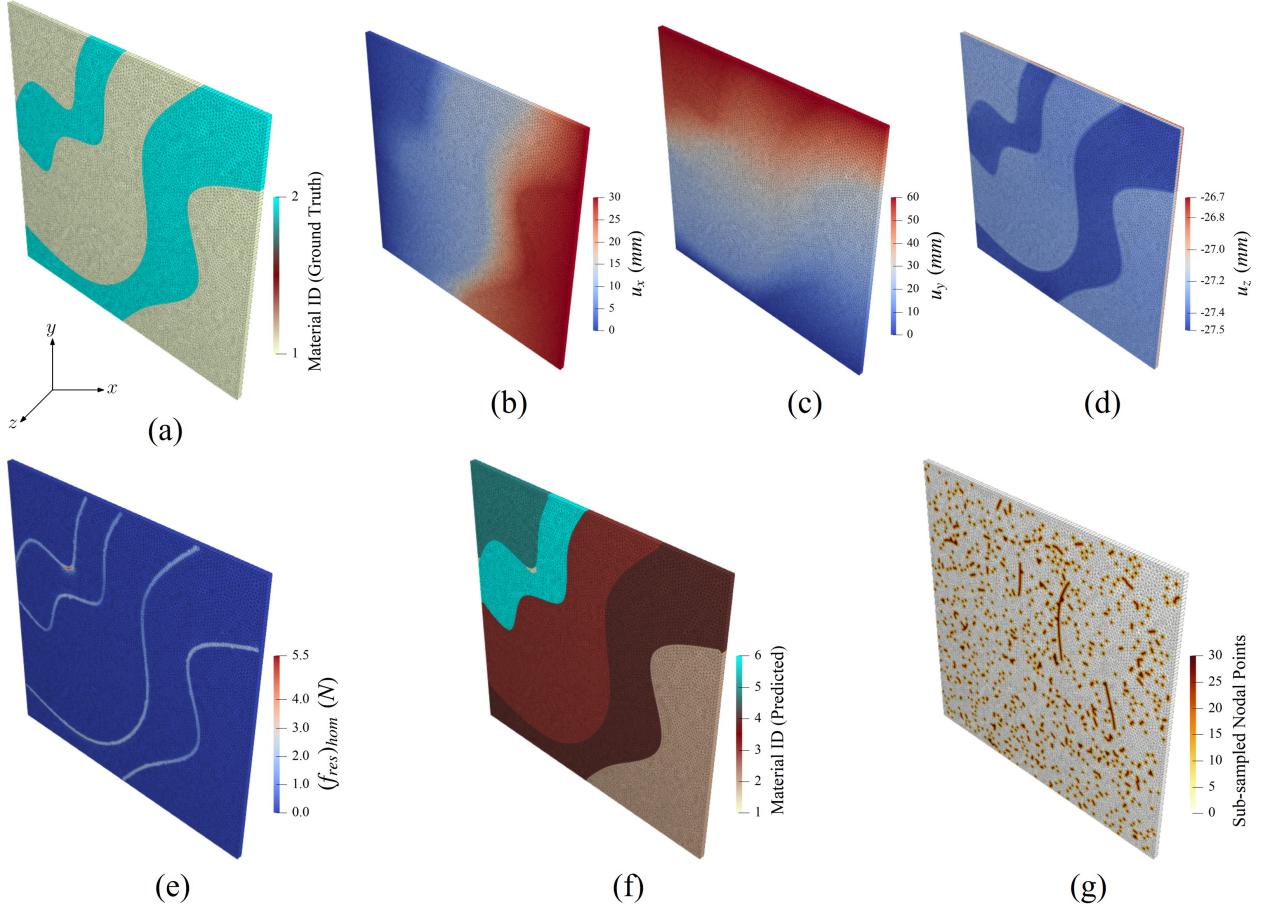


Pattern Type: Multiple inner inclusions | Mesh elements (C3D6): 10,816

Case	Material ID	Constitutive model	θ_1 (MPa)	θ_2 (MPa)	θ_3 (MPa)	θ_4 (MPa)	θ_5 (MPa)	θ_6 (MPa)	Free nodes sub-sampling
Ground Truth Refer (a)	1	ISH	4.00	0.50	0.30	—	—	21.00	—
	2	NH2 _a	5.40	—	—	—	—	15.00	
	3	HW	1.00	0.15	—	0.02	0.00	10.00	
	4	NH2 _b	1.80	—	—	—	—	6.00	
Moderate Noise $(\sigma_u = 2 \times 10^{-3})$	2	ISH	4.09 ± 0.07	0.51 ± 0.00	0.29 ± 0.02	0.00 ± 0.00	0.00 ± 0.00	21.12 ± 0.02	10%
	5	NH2 _a	5.34 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	14.82 ± 0.00	
	3	HW	0.67 ± 0.07	0.52 ± 0.03	0.01 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	9.63 ± 0.14	
	4	NH2 _b	0.06 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	2.81 ± 0.00	
High Noise $(\sigma_u = 5 \times 10^{-3})$ Refer (f)	2	ISH	4.91 ± 1.74	0.49 ± 0.06	0.17 ± 0.07	0.00 ± 0.00	0.00 ± 0.00	21.05 ± 0.13	10%
	5	NH2 _a	4.91 ± 0.26	0.00 ± 0.03	0.00 ± 0.03	0.00 ± 0.00	0.00 ± 0.00	13.77 ± 0.52	
	3	HW	1.40 ± 5.76	0.62 ± 0.09	0.00 ± 0.00	0.01 ± 0.01	0.02 ± 0.06	6.10 ± 0.76	
	4	NH2 _b	0.68 ± 0.51	0.05 ± 0.05	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	3.16 ± 0.78	

Table 7: True vs. predicted material parameters for multiple inner inclusions pattern with noise using Bayesian-EUCLID framework.

Figure 13: Model discovery for the multiple inner inclusions pattern (Figure 7c), with sub-figures (b)–(g) corresponding to the high noise case. (a) displays the ground truth material segment IDs. (b)–(d) show the displacement fields in the x , y , and z directions, respectively. (e) presents the residual force norm distribution (Equation 24) from the homogenized model. (f) illustrates the predicted material segments from the segmentation algorithm. (g) highlights the sub-sampled nodes from the D^{free} subset used to construct the heterogenized model, which is subsequently solved using the Bayesian framework.

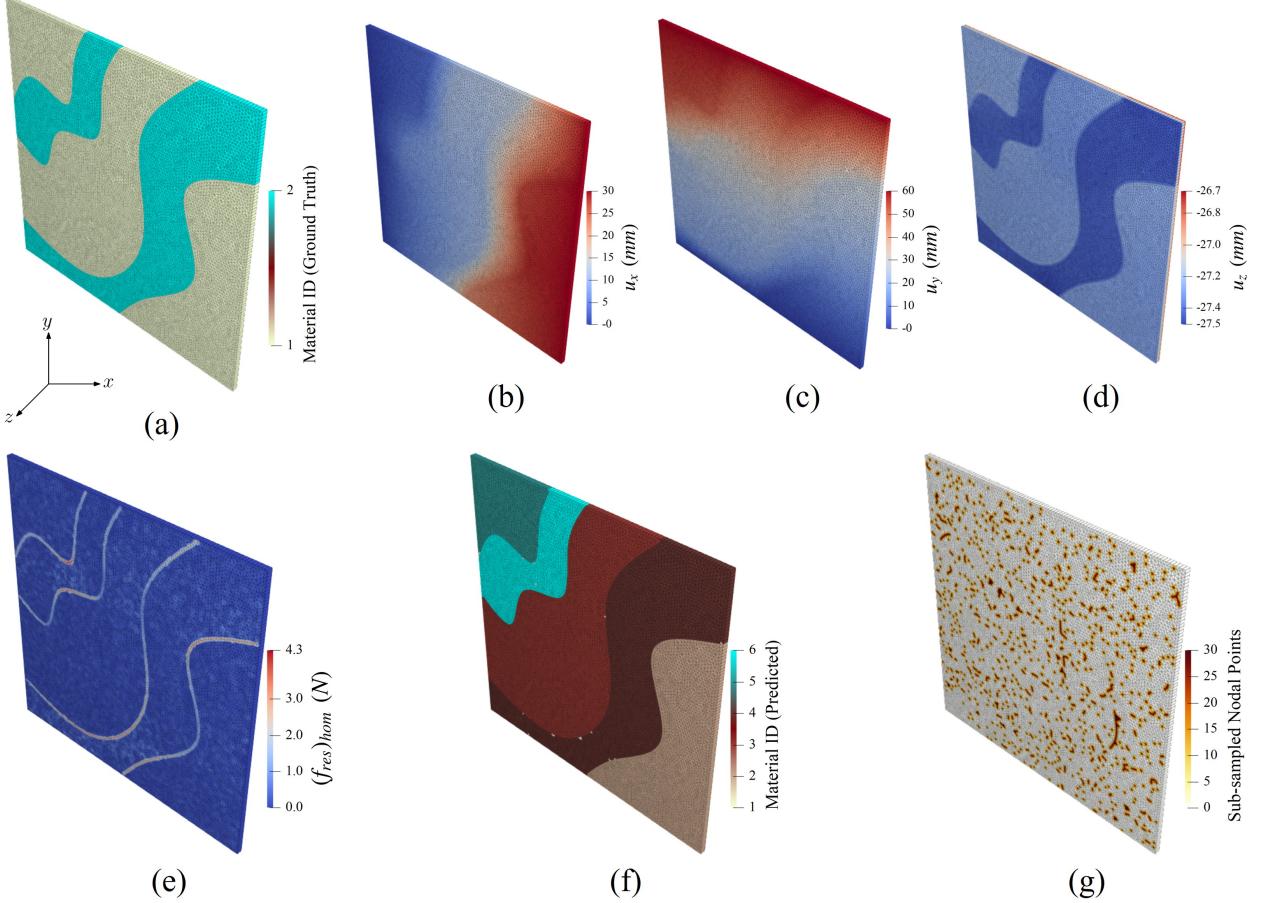


Pattern Type: Cahn-Hilliard | Mesh elements (C3D6): 21,040

Case	Material ID	Constitutive model	θ_1 (MPa)	θ_2 (MPa)	θ_3 (MPa)	θ_4 (MPa)	θ_5 (MPa)	θ_6 (MPa)	Free nodes sub-sampling
Ground Truth, (a)	1	NH2 _a	5.40	—	—	—	—	15.00	—
	2	NH2 _b	1.80	—	—	—	—	6.00	
No Noise Refer (f)	5	NH2 _a	5.40 ± 0.01	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	15.05 ± 0.03	
	6	NH2 _b	1.78 ± 0.00	0.00 ± 0.03	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	5.94 ± 0.01	
	3	NH2 _a	5.42 ± 0.01	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	15.12 ± 0.03	5%
	4	NH2 _b	1.80 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	6.00 ± 0.00	
	2	NH2 _a	5.42 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	15.02 ± 0.00	

Table 8: True vs. predicted material parameters for the Cahn-Hilliard pattern without noise using the Bayesian-EUCLID framework.

Figure 14: Model discovery for the Cahn-Hilliard pattern (Figure 7d), with sub-figures (b)–(g) corresponding to the zero noise case. (a) displays the ground truth material segment IDs. (b)–(d) show the displacement fields in the x , y , and z directions, respectively. (e) presents the residual force norm distribution (Equation 24) from the homogenized model. (f) illustrates the predicted material segments from the segmentation algorithm. (g) highlights the sub-sampled nodes from the $\mathcal{D}^{\text{free}}$ subset used to construct the heterogenized model, which is subsequently solved using the Bayesian framework.



Pattern Type: Cahn-Hilliard | Mesh elements (C3D6): 21,040

Case	Material ID	Constitutive model	θ_1 (MPa)	θ_2 (MPa)	θ_3 (MPa)	θ_4 (MPa)	θ_5 (MPa)	θ_6 (MPa)	Free nodes sub-sampling
Ground Truth, (a)	1	NH2 _a	5.40	—	—	—	—	15.00	—
	2	NH2 _b	1.80	—	—	—	—	6.00	
Moderate Noise ($\sigma_u = 2 \times 10^{-3}$)	5	NH2 _a	6.20 ± 0.09	0.00 ± 0.00	0.05 ± 0.05	0.00 ± 0.00	0.01 ± 0.01	16.76 ± 0.16	5%
	6	NH2 _b	1.82 ± 0.02	0.00 ± 0.03	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	6.03 ± 0.07	
	3	NH2 _a	5.28 ± 0.05	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	14.64 ± 0.12	
	4	NH2 _b	1.85 ± 0.01	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	6.12 ± 0.04	
	2	NH2 _a	4.44 ± 0.08	0.46 ± 0.02	0.07 ± 0.02	0.01 ± 0.01	0.00 ± 0.00	15.13 ± 0.07	
High Noise ($\sigma_u = 5 \times 10^{-3}$) Refer (f)	5	NH2 _a	6.74 ± 0.06	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	17.84 ± 0.20	10%
	6	NH2 _b	1.72 ± 0.03	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	5.71 ± 0.09	
	3	NH2 _a	5.03 ± 0.04	0.00 ± 0.00	0.03 ± 0.01	0.00 ± 0.00	0.00 ± 0.00	14.18 ± 0.11	
	4	NH2 _b	1.66 ± 0.02	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	5.52 ± 0.06	
	2	NH2 _a	5.39 ± 0.10	0.02 ± 0.03	0.02 ± 0.02	0.00 ± 0.00	0.00 ± 0.00	16.19 ± 0.11	

Table 9: True vs. predicted material parameters for the Cahn-Hilliard pattern with noise using the Bayesian-EUCLID framework.

Figure 15: Model discovery for the Cahn-Hilliard pattern (Figure 7d), with sub-figures (b)–(g) corresponding to the high noise case. (a) displays the ground truth material segment IDs. (b)–(d) show the displacement fields in the x , y , and z directions, respectively. (e) presents the residual force norm distribution (Equation 24) from the homogenized model. (f) illustrates the predicted material segments from the segmentation algorithm. (g) highlights the sub-sampled nodes from the $\mathcal{D}^{\text{free}}$ subset used to construct the heterogenized model, which is subsequently solved using the Bayesian framework.

Using the homogenized model formulation, a distinct imbalance in the residual force norm is observed along the complex Cahn–Hilliard pattern boundaries, as shown in [Figure 14e](#), [15e](#), and [S4e](#). Material region segmentation is performed using the threshold parameter $\lambda = 1.5$, resulting in the prediction of six material segments. The segmentation output is visualized in [Figure 14f](#), [15f](#), and [S4f](#), where segment ID 1 contains a small number of misclustered elements under both noise-free and noisy conditions. The remaining segments (IDs 2-6) exhibit well-defined and accurate material interfaces. Subsequently, the material parameter identification is carried out using 5% sub-sampling of the free nodes, and the noise-free results are reported in Table 8. Relative to the ground truth, material segments with IDs 2, 3, and 5 recover NH_{2a} model coefficients θ_1 and θ_6 close to 5.40 MPa and 15.00 MPa, respectively.

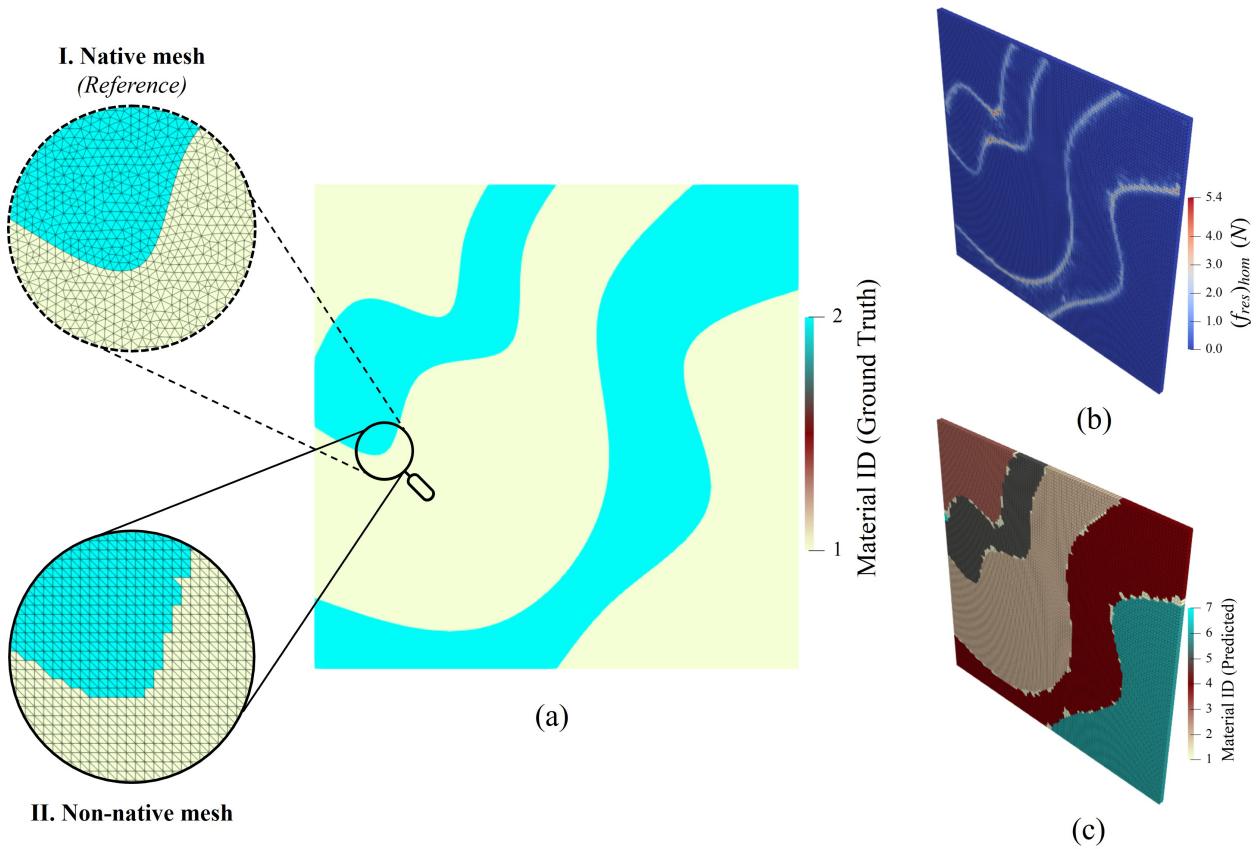
Similarly, the corresponding predictions of θ_1 and θ_6 for material segments 4 and 6 closely match their NH_{2b} ground truth values of 1.80 MPa and 6.00 MPa, respectively. All remaining material parameters θ_{2-5} are correctly predicted as zero with low standard deviation. A key advantage of the Hetero-EUCLID framework lies in its interpretability. While the segmentation algorithm identifies five major material regions, the predicted model parameters provide meaningful physical insight into material similarity. It can be inferred that segments 2, 3, and 5 exhibit nearly identical Neo-Hookean model coefficients, indicating that these segments represent the same material. Similarly, material regions with IDs 4 and 6 are identified as belonging to a second distinct material type. Thus, Hetero-EUCLID segments the domain and enables grouping regions based on interpretable material parameters, effectively identifying five spatial zones representing two unique material classes.

[Table 9](#) summarizes a comparative study of material parameter prediction performance for the Cahn–Hilliard pattern specimen under moderate and high noise settings. The resulting predictions of θ_1 and θ_6 across all identified Neo-Hookean regions lie within an acceptable tolerance of the ground truth and are associated with low standard deviations, indicating high confidence. Additionally, the remaining coefficients θ_{2-5} are correctly predicted as zero, promoting sparsity. [Supplementary Figure S8](#) demonstrates strong agreement between the strain energy density plots computed from the predicted model parameters and those from the corresponding ground truth. These results demonstrate the robustness and interpretability of the Hetero-EUCLID framework for complex heterogeneous configurations under high noise conditions.

3.3. Non-native mesh

In [subsection 3.2](#), we evaluated the performance of the Hetero-EUCLID framework across various case studies. For all configurations, the computational domain was discretized using wedge elements based on a known element connectivity matrix. Notably, the same mesh (referred to as the *native mesh*) used in the forward FEM simulations for generating displacement data was also used for the inverse formulation. In this section, we investigate further to emulate DIC experimental conditions more pragmatically, by evaluating the Hetero-EUCLID framework on a *non-native mesh*. A non-native mesh refers to a discretization where the element edges do not necessarily align with the underlying material interface boundaries. To investigate this scenario, we use the complex Type 4 Cahn–Hilliard heterogeneous pattern and discretize the domain using a uniform grid and generate wedge elements using Delaunay triangulation [59] on the surface, resulting in 19,602 uniformly distributed wedge elements. [Figure 16a](#) provides a comparative visualization between the native (I) and non-native (II) meshes, showing a zoomed-in view near a material interface. It can be observed that the native mesh aligns precisely with the material boundary, while the non-native mesh does not. In DIC experiments, the material interface geometry is inherently unknown during meshing; thus, this non-native mesh study is key to evaluating the robustness and practical applicability of Hetero-EUCLID framework under realistic experimental conditions.

Using the displacement field generated from ABAQUS/CAE for the Cahn–Hilliard pattern under non-equi-biaxial loading, we perform FE-based interpolation to compute full-field displacements at the nodes of the non-native mesh. As the FE-based interpolation is expected to have inherent noise, we do not add further Gaussian noise to the displacement data as was done in previous cases. [Figure 16b](#) illustrates the distribution of the residual force norm, where high magnitudes highlight the material interface.



Pattern Type: Cahn-Hilliard Non-native mesh elements (C3D6): 19,602									
Case	Material ID	Constitutive model	θ_1 (MPa)	θ_2 (MPa)	θ_3 (MPa)	θ_4 (MPa)	θ_5 (MPa)	θ_6 (MPa)	Free nodes sub-sampling
Ground Truth, (a)	1	NH2 _a	5.40	—	—	—	—	15.00	—
	2	NH2 _b	1.80	—	—	—	—	6.00	—
Using Ordinary Least Squares Regression	3	NH2 _a	7.58	0.29	0.74	-0.51	-0.30	21.14	
	5	NH2 _b	2.03	-0.15	-0.07	0.01	0.00	4.22	
	2	NH2 _a	4.95	-0.36	-0.19	0.13	0.01	12.12	—
	4	NH2 _b	2.25	-0.24	-0.07	0.01	0.00	3.68	
	6	NH2 _a	8.98	-1.65	-0.49	0.58	-0.06	18.88	
Using Bayesian framework Refer (c)	3	NH2 _a	5.94 ± 0.27	0.11 ± 0.12	0.00 ± 0.00	0.01 ± 0.04	0.00 ± 0.00	17.20 ± 0.61	
	5	NH2 _b	1.31 ± 0.03	0.02 ± 0.01	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	4.39 ± 0.06	
	2	NH2 _a	4.84 ± 0.25	0.20 ± 0.03	0.01 ± 0.01	0.01 ± 0.00	0.00 ± 0.00	14.26 ± 0.69	10%
	4	NH2 _b	1.76 ± 0.01	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	5.64 ± 0.02	
	6	NH2 _a	5.89 ± 0.04	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.01	15.26 ± 0.02	

Table 10: True vs. predicted material parameters for the Cahn-Hilliard pattern using *non-native* meshing.

Figure 16: Model discovery for the Cahn–Hilliard pattern (Figure 7d) using *non-native* meshing. (a) shows the ground truth material segment IDs with visual comparison of (I) native and (II) non-native mesh views. (b) presents the residual force norm distribution (Equation 24) from the homogenized model. (c) illustrates the predicted material segments obtained through the segmentation algorithm. Table 10 summarizes the ground truth and predicted material parameters obtained by solving the heterogenized model using the ordinary least squares regression and Bayesian framework.

Subsequently, the segmentation is performed using the random seed-based island growth algorithm, and

the resulting material segments are shown in [Figure 16c](#). While the algorithm successfully identifies five major material zones, the interfaces appear less sharp, which is expected and justified given the non-aligned structure of the non-native mesh grid compared to those observed with the native mesh configuration.

[Table 10](#) compares material parameter predictions using ordinary least squares (OLS) regression and the Bayesian framework. When the heterogeneous model (refer to [Equation 20](#)) is solved using OLS, the predicted NH2 model coefficients ($\theta_{1,6}$)_{OLS} exhibit notable deviations from the ground truth. Moreover, the intermediate model parameters (θ_{2-5})_{OLS} yield negative values, which may render the strain energy density non-polyconvex, and reflect overfitting under noise. In contrast, the Bayesian framework enforces positivity on the material parameters $\boldsymbol{\theta}$ via truncated normal distribution priors, ensuring physical admissibility. The material parameters predictions via the Bayesian approach are closer to the ground truth and are associated with lower standard deviations, indicating a higher level of statistical confidence. Additionally, the remaining coefficients θ_{2-5} are correctly predicted as zero, promoting sparsity. This high confidence is further corroborated by the consistently high R^2 scores observed across the subplots of strain energy density shown in [Supplementary Figure S9](#). Furthermore, the interpretability of the Hetero-EUCLID method enables physically intuitive grouping of the predicted segments: regions labeled 2, 3, and 6 are identified as having identical material properties, while segments 4 and 5 correspond to a second distinct hyperelastic material. These results demonstrate that Hetero-EUCLID remains accurate, interpretable, and robust under experimental constraints such as non-native meshing for inverse material parameter identification problems.

Specimen Type	R_{\max} [N] (Maximum boundary force)	Noise Level	λ (Tunable scalar)	$\mu_{f_{\text{res}}}$ [N] (Mean)	$\sigma_{f_{\text{res}}}$ [N] (Standard deviation)	$\frac{\mu_{f_{\text{res}}}}{\sigma_{f_{\text{res}}}}$	$\frac{\sigma_{f_{\text{res}}}}{R_{\max}}$
Type 0 (Homogeneous)	904.56	Zero	—	2.38×10^{-5}	1.48×10^{-5}	1.608	1.64×10^{-8}
		Moderate	—	0.122	0.074	1.664	8.13×10^{-5}
		High	—	0.247	0.135	1.826	1.49×10^{-4}
Type 1 (Figure 7a)	344.44	Zero	2.00	0.042	0.252	0.168	7.31×10^{-4}
		Moderate	2.00	0.085	0.247	0.343	7.17×10^{-4}
		High	1.50	0.134	0.243	0.552	7.04×10^{-4}
Type 2 (Figure 7b)	879.28	Zero	1.50	0.065	0.380	0.172	4.32×10^{-4}
		Moderate	1.50	0.193	0.371	0.520	4.22×10^{-4}
		High	2.12	0.328	0.377	0.871	4.28×10^{-4}
Type 3 (Figure 7c)	969.58	Zero	1.70	0.185	0.906	0.204	9.34×10^{-4}
		Moderate	1.70	0.338	0.902	0.375	9.30×10^{-4}
		High	1.80	0.507	0.888	0.571	9.16×10^{-4}
Type 4 (Figure 7d)	562.26	Zero	1.50	0.095	0.472	0.202	8.40×10^{-4}
		Moderate	1.50	0.186	0.469	0.396	8.34×10^{-4}
		High	1.50	0.272	0.464	0.587	8.25×10^{-4}

Table 11: Statistical parameters (mean and standard deviation) and ratios for different specimen types and noise levels.

3.4. A comment on detecting noise

[Table 11](#) provides statistical details of the 2D nodal residual force f_{res}^a corresponding to each type of heterogeneity and noise-level considered in the study. Additionally, we provide statistical details of residual forces in a homogeneous specimen with the same thin square plate domain geometry in the same table for comparison. It is evident from the results that the dimensionless ratio: $\left(\frac{\mu_{f_{\text{res}}}}{\sigma_{f_{\text{res}}}}\right)$ is a measure of noise in the displacement data, with a higher ratio indicating a higher displacement noise or a lack of heterogeneity in the domain. Another important observation is that alongside the previous ratio, the dimensionless ratio $\left(\frac{\sigma_{f_{\text{res}}}}{R_{\max}}\right)$ provides an indication of the existence of heterogeneity in the domain. If $\left(\frac{\sigma_{f_{\text{res}}}}{R_{\max}}\right) \leq 10^{-5}$ and

$\left(\frac{\mu_{f_{\text{res}}}}{\sigma_{f_{\text{res}}}}\right) \geq 1$, it is reasonable to consider the domain to be nominally homogeneous. However, further increase in displacement noise can be expected to increase both ratios and make identifying heterogeneities (or the lack thereof) in the domain challenging. The statistical trends observed from f_{res}^a are thus useful to characterize displacement noise and the existence of heterogeneities from actual experimental data.

4. Conclusion and outlook

In this study, we presented and validated Hetero-EUCLID, a physics-constrained, interpretable framework for discovering the segmentation and region-wise constitutive models in heterogeneous hyperelastic materials under 3D plane stress conditions. We demonstrated that Hetero-EUCLID accurately identifies both the material interface boundaries and the underlying constitutive laws for each region for different heterogeneity patterns, including cross-inclusion, split-domains, non-circular inclusions, and Cahn–Hilliard configurations. The Bayesian framework provided sparsity and a parsimonious model parameter under quantified uncertainty, enhancing confidence in predictions. Its performance and efficacy have been validated through empirical checks across various noise levels and non-native mesh discretizations, confirming the robustness and interpretability of the proposed approach. The computations were performed on a machine with 12th Gen Intel® Core™ i9-12900K processor having clock frequencies of 3.2 GHz and 32 GB RAM. We employed 3 cores in parallel, each corresponding to a parallel Markov chain. The average compute time was seen to be around 1 hr 20 mins, with lower noise data requiring significantly lower compute time. The compute time can be further improved by using more number of shorter parallel chains. We did not include anisotropy in each hyperelastic material segments in this work for the sake of brevity in the feature library. However, anisotropic hyperelastic materials can be easily accounted for by adding directional stretch invariant-based energy terms in the feature library [46].

As part of future work, we aim to extend Hetero-EUCLID to accommodate more complex history-dependent constitutive behaviors such as viscoelasticity to model a broader range of soft tissues and composites. As mentioned in Shi et al. [45], the paradigm of first segmenting the material and then identifying the constitutive parameters of each segment is highly dependent on the ability to have clean inter-material segments. However, this is not the case in situations involving functionally graded materials, and most practical experiments where the material segments have smooth (not sudden) transitions between properties. Furthermore, in the current study, we show that a high enough displacement noise would undermine the accuracy of the material segmentation, further reducing the accuracy of predicting each material's properties. We therefore intend to explore frameworks that provide a position-dependent field of the material properties, as opposed to segment-specific properties. An experimental validation using DIC/ Digital Volume Correlation (DVC) data will also provide further insights into the merits, practical limitations, and improvements necessary in the proposed framework.

Declaration of competing interest

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

Code and data availability

The codes and data generated during the current study are available from the authors upon reasonable request.

Acknowledgments

AJ would like to acknowledge funding by Anusandhan National Research Foundation (ANRF) through Grant No. ANRF/ECRG/2024/005303/ENS.

Appendix A. Overview of Bayesian-EUCLID framework

Appendix A.1. Parsimony-driven Model Prior

We adopt a probabilistic framework that simultaneously performs model selection and quantifies uncertainty to introduce sparsity and maintain interpretability in the discovered model. Instead of relying on conventional deterministic sparse regression, we use a hierarchical Bayesian formulation with spike–slab priors, initially introduced by Nayek et al. [54] in a supervised setting, and later adapted to an unsupervised formulation by Joshi et al. [46]. Spike–slab priors are characterized by a sharp peak (“spike”) at zero and broad, shallow tails (“slab”). A random variable may be drawn from either the spike or slab distribution. Sampling from the spike yields values at zero, reflecting our bias toward parsimonious (sparse) models, whereas sampling from the slab allows non-zero values with a near-uniform spread.

For the model parameter vector $\boldsymbol{\theta}$, a binary indicator vector $\mathbf{z} \in [0, 1]^{n_f}$ is introduced, along with random variables $\nu_s \geq 0$ and $\sigma^2 \geq 0$. The conditional prior for each component is

$$p(\theta_i | z_i, \nu_s, \sigma^2) = \begin{cases} p_{\text{spike}}(\theta_i) = \delta(\theta_i) & \text{if } z_i = 0, \\ p_{\text{slab}}(\theta_i | \nu_s, \sigma^2) = \mathcal{N}_+(0, \nu_s \sigma^2) & \text{if } z_i = 1, \end{cases} \quad \forall i = 1, \dots, n_f \quad (\text{A.1})$$

where \mathcal{N}_+ denotes a non-negative truncated normal distribution with zero mean and variance $\nu_s \sigma^2$ (slab). The truncated normal, \mathcal{N}_+ is obtained from the standard normal distribution by restricting samples to non-negative values ($\theta_i \geq 0$), a requirement that ensures physical admissibility. When $z_i = 0$, the i^{th} feature is considered inactive, meaning $\theta_i = 0$, and its probability density reduces to a Dirac delta function (spike) centered at zero. Conversely, if $z_i = 1$, the feature is active and θ_i is drawn from a truncated normal distribution \mathcal{N}_+ . Assuming that the elements of $\boldsymbol{\theta}$ are independent and identically distributed (i.i.d.), the joint prior for $\boldsymbol{\theta}$ can be expressed as

$$p(\boldsymbol{\theta} | \mathbf{z}, \nu_s, \sigma^2) = p_{\text{slab}}(\boldsymbol{\theta}_r | \nu_s, \sigma^2) \prod_{i:z_i=0} p_{\text{spike}}(\theta_i) \quad \text{with} \quad p_{\text{slab}}(\boldsymbol{\theta}_r | \nu_s, \sigma^2) = \mathcal{N}_+(\mathbf{0}, \sigma^2 \nu_s \mathbf{I}_r) \quad (\text{A.2})$$

where, $\boldsymbol{\theta}_r \in \mathbb{R}_+^r$ denotes the reduced vector having components of $\boldsymbol{\theta}$ corresponds to only the slab, or active features and \mathbf{I}_r is the $r \times r$ identity matrix. Sampling from the multivariate truncated normal distribution is performed using the algorithm proposed by Botev [60]. The binary variables z_i are drawn from a Bernoulli distribution with activation probability p_0 as follows

$$z_i | \sim \text{Bern}(p_0), \quad \text{where} \quad p_0 \in [0, 1] \quad (\text{A.3})$$

The hyper-priors for the random variables (ν_s, σ^2, p_0) are modeled as

$$\nu_s \sim \mathcal{IG}(a_\nu, b_\nu), \quad \sigma^2 \sim \mathcal{IG}(a_\sigma, b_\sigma), \quad p_0 \sim \text{Beta}(a_p, b_p) \quad (\text{A.4})$$

where \mathcal{IG} and Beta represent the Inverse gamma and Beta distribution, respectively, and (a, b) denotes the set of hyperparameters for the random variables ν_s , σ^2 , and p_0 .

Appendix A.2. Model Identification through Posterior Estimation

The Bayesian learning task now reduces to estimating the joint posterior probability distribution of the random variables $\{\boldsymbol{\theta}, \mathbf{z}, p_0, \nu_s, \sigma^2\}$ conditioned on the observed data \mathbf{A} and \mathbf{b} and the weak form of momentum balance. Applying Bayes’ theorem, the posterior can be expressed as:

$$p(\boldsymbol{\theta}, \mathbf{z}, p_0, \nu_s, \sigma^2 | \mathbf{A}, \mathbf{b}) \propto \underbrace{p(\mathbf{b} | \boldsymbol{\theta}, \mathbf{z}, p_0, \nu_s, \sigma^2, \mathbf{A})}_{\text{physics-constrained likelihood}} \underbrace{p(\boldsymbol{\theta}, \mathbf{z}, p_0, \nu_s, \sigma^2)}_{\text{spike-slab model prior}} \quad (\text{A.5})$$

It is important to note that the direct analytical sampling from Equation A.5 is challenging due to the presence of spike-slab priors, therefore Markov Chain Monte Carlo (MCMC) methods is deployed. In this work, we adopt a Gibbs sampling strategy [61] for posterior distribution estimation. The Gibbs sampler

requires the conditional posterior distributions of all random variables. Closed-form expressions for these conditional posteriors were derived by Nayek et al. [54], and a concise summary of these results is provided in [Table A.2](#):

Parameter/Variable	Conditional posterior distribution
$\boldsymbol{\theta}_r$	$\boldsymbol{\theta}_r \mid z, p_0, \nu_s, \sigma^2, \mathbf{A}, \mathbf{b} \sim \mathcal{N}_+(\boldsymbol{\mu}, \sigma^2 \boldsymbol{\Sigma})$ $\boldsymbol{\Sigma} = (\mathbf{A}_r^T \mathbf{A}_r + \nu_s^{-1} \mathbf{I}_r)^{-1}, \quad \boldsymbol{\mu} = \boldsymbol{\Sigma} \mathbf{A}_r^T \mathbf{b}$
σ^2	$\sigma^2 \mid \boldsymbol{\theta}, z, p_0, \nu_s, \mathbf{A}, \mathbf{b} \sim \mathcal{IG}\left(a_\sigma + \frac{N}{2}, b_\sigma + \frac{1}{2}(\mathbf{b}^T \mathbf{b} - \boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu})\right)$
ν_s	$\nu_s \mid \boldsymbol{\theta}, z, p_0, \sigma^2, \mathbf{A}, \mathbf{b} \sim \mathcal{IG}\left(a_\nu + \frac{s_z}{2}, b_\nu + \frac{\boldsymbol{\theta}_r^T \boldsymbol{\theta}_r}{2\sigma^2}\right)$
p_0	$p_0 \mid \boldsymbol{\theta}, z, \nu_s, \sigma^2, \mathbf{A}, \mathbf{b} \sim \text{Beta}(a_p + s_z, b_p + n_f - s_z)$
z_i	$z_i \mid \boldsymbol{\theta}, \mathbf{z}_{-i}, p_0, \nu_s, \sigma^2, \mathbf{A}, \mathbf{b} \sim \text{Bern}(\xi_i)$ $\xi_i = p_0 \left[p_0 + \frac{p(\mathbf{b} \mid z_i = 0, \mathbf{z}_{-i}, \nu_s, \mathbf{A})}{p(\mathbf{b} \mid z_i = 1, \mathbf{z}_{-i}, \nu_s, \mathbf{A})} (1 - p_0) \right]^{-1}$

Table A.2: Full conditional posterior distributions for model parameters.

The matrix $\mathbf{A}_r \in \mathbb{R}^{N \times s_z}$ is constructed by concatenating only those columns $i = 1, \dots, n_f$ of \mathbf{A} for which $z_i = 1$; hence, the distribution applies exclusively and jointly to the active features $\boldsymbol{\theta}_r$. Here, \mathbf{I}_r denotes the $s_z \times s_z$ identity matrix, and $s_z = \sum_{i=1}^{n_f} z_i$ is the total number of active features. For any $z_i = 0$, the associated θ_i is set to zero. In our framework following Joshi et al. [46], however, physical admissibility requires $\boldsymbol{\theta} \geq 0$, which modifies the conditional posterior on $\boldsymbol{\theta}$ from an unconstrained multivariate normal distribution to a constrained (non-negative) multivariate normal \mathcal{N}^+ , while leaving the other conditional distributions unchanged as formulated by Nayek et al. [54].

At each iteration of the Gibbs sampler, the components of \mathbf{z} are updated in a random sequence. The notation \mathbf{z}_{-i} refers to the vector \mathbf{z} with its i -th component z_i removed. The marginal likelihood $p(\mathbf{b} \mid \mathbf{z}, \nu_s, \mathbf{A})$ is then obtained by integrating out $\boldsymbol{\theta}$ and σ^2 from the likelihood in [Equation 22](#), and is expressed as

$$p(\mathbf{b} \mid \mathbf{z}, \nu_s, \mathbf{A}) = \frac{\Gamma(a_\sigma + 0.5N)}{(2\pi)^{N/2} (\nu_s)^{s_z/2} \Gamma(a_\sigma)} \left[\det \left((\mathbf{A}_r^T \mathbf{A}_r + \nu_s^{-1} \mathbf{I}_r)^{-1} \right) \right]^{1/2} \times \left[b_\sigma + 0.5 \mathbf{b}^T \left(\mathbf{I}_N - \mathbf{A}_r \left(\mathbf{A}_r^T \mathbf{A}_r + \nu_s^{-1} \mathbf{I}_r \right)^{-1} \mathbf{A}_r^T \right) \mathbf{b} \right]^{-(a_\sigma + 0.5N)}. \quad (\text{A.6})$$

where $\Gamma(\cdot)$ is the Gamma function, N is the number of observations, and \mathbf{A}_r corresponds to the active feature sub-matrix defined by the current state of \mathbf{z} .

In the Gibbs sampling process, each variable in the set $\{\boldsymbol{\theta}, z, p_0, \nu_s, \sigma^2\}$ is iteratively updated by drawing from its corresponding conditional posterior distribution, as defined in [Table A.2](#). This sequential updating generates a Markov chain of length N_G , which serves as an empirical approximation of the joint posterior $p(\boldsymbol{\theta}, z, p_0, \nu_s, \sigma^2 \mid \mathbf{A}, \mathbf{b})$

$$\boldsymbol{\theta}^{(0)} \rightarrow \sigma^{2(0)} \rightarrow \nu_s^{(0)} \rightarrow p_0^{(0)} \rightarrow \mathbf{z}^{(0)} \rightarrow \dots \rightarrow \boldsymbol{\theta}^{(N_G)} \rightarrow \sigma^{2(N_G)} \rightarrow \nu_s^{(N_G)} \rightarrow p_0^{(N_G)} \rightarrow \mathbf{z}^{(N_G)} \quad (\text{A.7})$$

To ensure the chain reflects the target distribution, the initial N_{burn} samples are discarded as burn-in, thereby mitigating the influence of potentially poor starting states. Additionally, N_{chains} independent Markov chains are generated, each initialized with a different randomly chosen starting point. After burn-in removal, these independent chains are concatenated to produce the final sample set used for posterior inference.

Appendix B. Optimal sub-sampling criterion at material interface boundaries

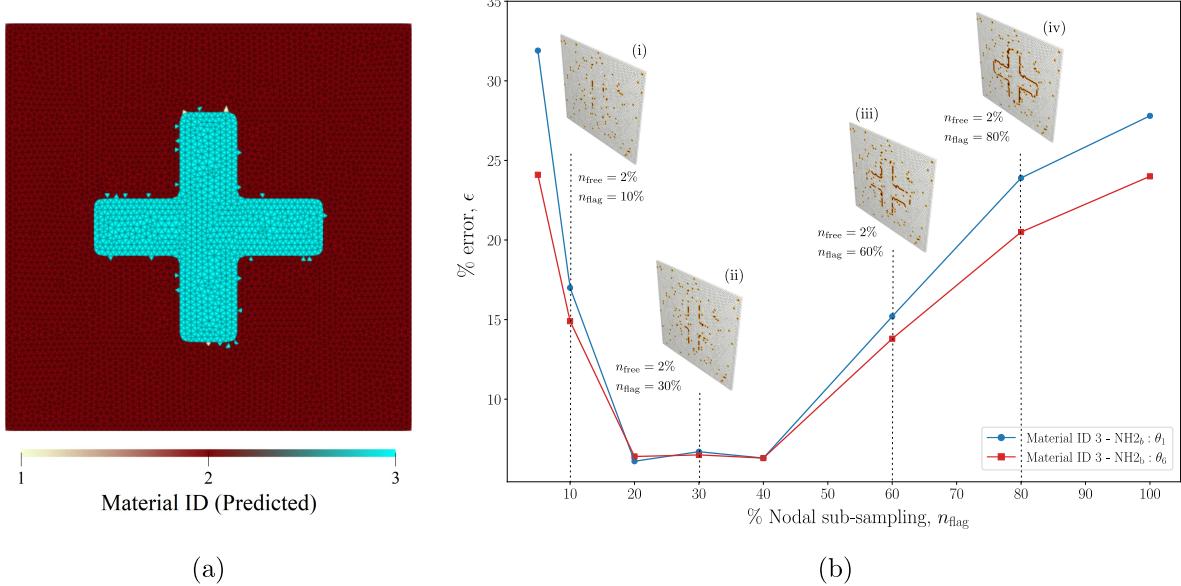


Figure B.17: Study on the effect of node sub-sampling on misclustered inter-segment boundaries for Type 1 heterogeneous domain.

To assess the impact of interface node sub-sampling on the robustness of Hetero-EUCLID model predictions under high noise conditions ($\sigma_{u, \text{high}} = 5 \times 10^{-3}$), we conducted a parameter study by varying the subset of sampled flagged nodes and evaluating the percentage error ϵ , between the predicted and ground truth material parameters. Specifically, we define n_{flag} as a percentage of all flagged nodes N_{flag} , selected by sorting the entries of residual force norm vector $(f_{\text{res}}^a)_{\text{het}}$, in ascending order and choosing the lowest $n_{\text{flag}}\%$ of the sorted column vector.

$$\epsilon_k = \left| \frac{\theta_k^{\text{predicted}} - \theta_k^{\text{true}}}{\theta_k^{\text{true}}} \right| \times 100 \%, \quad \text{where } k = 1, \dots, n_f \quad (\text{B.1})$$

For the Type 1: Cross-shaped inclusion pattern (see Figure 7a) governed by the NH2–NH2 model, we introduced a random misclustering of wedge elements near the material interface (see Figure B.17a) to simulate a realistic segmentation error. Figure B.17b shows how the percentage error in predicted material parameters (θ_1 and θ_6) for cross inclusion changes as n_{flag} increases from 5% to 100%, while the free node sub-sampling rate n_{free} is kept constant at 2%. A characteristic valley-shaped trend indicates that an intermediate sub-sampling level performs best under high noise conditions. The corresponding spatial distribution of the sampled interface nodes for 10%, 30%, 60%, and 80% is visualized in Figure B.17b(i)–(iv). It is evident that selecting 20%–40% of the lowest residual force nodes leads to more accurate and stable parameter predictions, indicating the robustness of the Hetero-EUCLID framework. However, higher values of n_{flag} do not significantly improve accuracy and come with increased computational cost. Hence, we chose $n_{\text{flag}} = 20\%$ as the optimal setting for interface node sub-sampling throughout the current study.

Appendix C. Pseudo-code for heterogeneous hyperelastic material model discovery

Algorithm 2 Hetero-EUCLID algorithm

Input: Full-field displacement data \mathbf{u} , boundary reaction forces \mathbf{R}

- 1: $\mathcal{S} \leftarrow$ homogeneous material model ▷ modeling assumption
- 2: **for** $t = 1, \dots, n_t$ **do** ▷ n_t : number of load steps
- 3: $\theta^{\text{ols}} \leftarrow \text{ols}(A, b)$ ▷ Solve $\min_{\theta} \|A\theta - b\|^2$
- 4: $(f_{\text{res}})_{\text{hom}} \leftarrow A_{\text{free}}\theta^{\text{ols}} - b_{\text{free}}$ ▷ Vector of (f_x^a, f_y^a, f_z^a) for all free nodes
- 5: **for** each free node $a \in \mathcal{D}^{\text{free}}$ **do**
- 6: $f_{\text{res}}^a \leftarrow \sqrt{(f_x^a)^2 + (f_y^a)^2}$ ▷ Euclidean Norm
- 7: **end for**
- 8: **end for**
- 9: $\{S_k : k \in 1 \dots (i-1)\} \leftarrow \text{getMaterialSegments}(\mathbf{N}_{\text{flag}}, \text{Nodal positions}, \text{Connectivity})$ ▷ Algorithm 1
(refer Appendix A)
- 10: **Bayesian EUCLID:**
- 11: $n_{\text{free}} \leftarrow \text{SampleFreeNodes}(\mathcal{D}^{\text{free}}, 0.02-0.10)$
- 12: $n_{\text{flag}} \leftarrow \text{SampleInterfaceNodes}(\mathcal{D}^{\text{interface}}, 0.20-0.40, (f_{\text{res}}^a)_{\text{het}})$ (refer Appendix B)
- 13: Posterior inference performed via Markov Chain Monte Carlo (MCMC)
- 14: **for** $k = 1, \dots, N_{\text{chains}}$ **do**
- 15: Allocate chain storage C_k of size $(N_{\text{burn}} + N_G)$
- 16: Initialize $\theta_r \leftarrow$ model coefficients
- 17: Initialize $\sigma^2 \leftarrow$ noise variance
- 18: Initialize $v_s \leftarrow$ scaling hyperparameter
- 19: Initialize $p_0 \leftarrow$ prior weight
- 20: Initialize $z \leftarrow$ latent node assignments
- 21: **for** $q = 1, \dots, (N_{\text{burn}} + N_G)$ **do** ▷ Random permutation of node indices
- 22: Sample $\theta \sim p(\theta | \cdot)$
- 23: Sample $\sigma^2, v_s, p_0 \sim p(\sigma^2, v_s, p_0 | \cdot)$
- 24: $\text{order} \leftarrow \text{Perm}(1, \dots, n_f)$
- 25: **for** $j = 1, \dots, n_f$ **do**
- 26: $i \leftarrow \text{order}[j]$
- 27: Sample $z_i \sim p(z_i | \theta, \sigma^2, v_s, p_0)$
- 28: **end for**
- 29: Record $C_k[q] \leftarrow \{\theta, \sigma^2, v_s, p_0, z\}$
- 30: **end for**
- 31: **end for**
- 32: Discard first N_{burn} samples from each chain
- 33: Retain last N_G samples as approximate posterior draws
- 34: **Output:** Material parameters $(\boldsymbol{\theta}^{(1)}, \dots, \boldsymbol{\theta}^{(k)})$ for each segment, with standard deviations and mean

References

- [1] Y. Guo, M. R. Mofrad, A. B. Tepole, On modeling the multiscale mechanobiology of soft tissues: Challenges and progress, *Biophysics Reviews* 3 (2022).
- [2] J. Wei, F. Pan, H. Ping, K. Yang, Y. Wang, Q. Wang, Z. Fu, Bioinspired additive manufacturing of hierarchical materials: From biostructures to functions, *Research* 6 (2023) 0164.
- [3] D. Lee, W. Chen, L. Wang, Y.-C. Chan, W. Chen, Data-driven design for metamaterials and multiscale systems: a review, *Advanced Materials* 36 (2024) 2305254.

- [4] N. S. Burris, Z. Bian, J. Dominic, J. Zhong, I. B. Houben, T. M. van Bakel, H. J. Patel, B. D. Ross, G. E. Christensen, C. R. Hatt, Vascular deformation mapping for ct surveillance of thoracic aortic aneurysm growth, *Radiology* 302 (2022) 218–225.
- [5] S. Lin, M. C. Morgant, D. M. Marín-Castrillón, P. M. Walker, L. S. A. Glélé, A. Boucher, B. Presles, O. Bouchot, A. Lalande, Aortic local biomechanical properties in ascending aortic aneurysms, *Acta Biomaterialia* 149 (2022) 40–50.
- [6] K. Vander Linden, E. Vanderveken, L. Van Hoof, L. Maes, H. Fehervary, S. Dreesen, A. Hendrickx, P. Verbrugghe, F. Rega, B. Meuris, et al., Stiffness matters: improved failure risk assessment of ascending thoracic aortic aneurysms, *JTCVS open* 16 (2023) 66–83.
- [7] A. Goriely, *The mathematics and mechanics of biological growth*, volume 45, Springer, 2017.
- [8] A. Micalet, E. Moeendarbary, U. Cheema, 3d in vitro models for investigating the role of stiffness in cancer invasion, *ACS Biomaterials Science & Engineering* 9 (2021) 3729–3741.
- [9] F. Hild, S. Roux, Digital image correlation: from displacement measurement to identification of elastic properties—a review, *Strain* 42 (2006) 69–80.
- [10] T. Marwala, *Finite element model updating using computational intelligence techniques: applications to structural dynamics*, Springer Science & Business Media, 2010.
- [11] M. Grediac, F. Pierron, S. Avril, E. Toussaint, The virtual fields method for extracting constitutive parameters from full-field measurements: a review, *Strain* 42 (2006) 233–253.
- [12] T. Kirchdoerfer, M. Ortiz, Data-driven computational mechanics, *Computer Methods in Applied Mechanics and Engineering* 304 (2016) 81–101.
- [13] R. Ibañez, D. Borzacchiello, J. V. Aguado, E. Abisset-Chavanne, E. Cueto, P. Ladeveze, F. Chinesta, Data-driven non-linear elasticity: constitutive manifold construction and problem discretization, *Computational Mechanics* 60 (2017) 813–826.
- [14] T. Kirchdoerfer, M. Ortiz, Data-driven computing in dynamics, *International Journal for Numerical Methods in Engineering* 113 (2018) 1697–1710.
- [15] S. Conti, S. Müller, M. Ortiz, Data-driven problems in elasticity, *Archive for Rational Mechanics and Analysis* 229 (2018) 79–123.
- [16] L. T. K. Nguyen, M.-A. Keip, A data-driven approach to nonlinear elasticity, *Computers & Structures* 194 (2018) 97–115.
- [17] R. Eggersmann, T. Kirchdoerfer, S. Reese, L. Stainier, M. Ortiz, Model-free data-driven inelasticity, *Computer Methods in Applied Mechanics and Engineering* 350 (2019) 81–99.
- [18] P. Carrara, L. De Lorenzis, L. Stainier, M. Ortiz, Data-driven fracture mechanics, *Computer Methods in Applied Mechanics and Engineering* 372 (2020) 113390.
- [19] K. Karapiperis, M. Ortiz, J. E. Andrade, Data-driven nonlocal mechanics: Discovering the internal length scales of materials, *Computer Methods in Applied Mechanics and Engineering* 386 (2021) 114039.
- [20] I. B. Rocha, P. Kerfriden, F. van Der Meer, On-the-fly construction of surrogate constitutive models for concurrent multiscale mechanical analysis through probabilistic machine learning, *arXiv preprint arXiv:2007.07749* (2020).
- [21] J. N. Fuhr, M. Marino, N. Bouklas, Local approximate gaussian process regression for data-driven constitutive models: development and comparison with neural networks, *Computer Methods in Applied Mechanics and Engineering* 388 (2022) 114217.

- [22] S.-H. Park, E.-H. Lee, H. Choi, J. Y. Lee, A thermal–elastic–plastic constitutive model using the radial basis function neural network and application for an energy efficient warm forming process, International Journal of Precision Engineering and Manufacturing 22 (2021) 329–343.
- [23] J. Crespo, M. Latorre, F. J. Montáns, Wypiwyg hyperelasticity for isotropic, compressible materials, Computational Mechanics 59 (2017) 73–92.
- [24] T. Sussman, K.-J. Bathe, A model of incompressible isotropic hyperelastic material behavior using spline interpolations of tension–compression test data, Communications in numerical methods in engineering 25 (2009) 53–63.
- [25] J. Ghaboussi, J. Garrett Jr, X. Wu, Knowledge-based modeling of material behavior with neural networks, Journal of engineering mechanics 117 (1991) 132–153.
- [26] M. Fernández, M. Jamshidian, T. Böhlke, K. Kersting, O. Weeger, Anisotropic hyperelastic constitutive models for finite deformations combining material theory and data-driven approaches with application to cubic lattice metamaterials, Computational Mechanics 67 (2021) 653–677.
- [27] D. K. Klein, M. Fernández, R. J. Martin, P. Neff, O. Weeger, Polyconvex anisotropic hyperelasticity with neural networks, Journal of the Mechanics and Physics of Solids 159 (2022) 104703.
- [28] N. N. Vlassis, W. Sun, Sobolev training of thermodynamic-informed neural networks for interpretable elasto-plasticity models with level set hardening, Computer Methods in Applied Mechanics and Engineering 377 (2021) 113695.
- [29] S. Kumar, S. Tan, L. Zheng, D. M. Kochmann, Inverse-designed spinodoid metamaterials, npj Computational Materials 6 (2020) 73.
- [30] J.-H. Basteck, S. Kumar, B. Telgen, R. N. Glaesener, D. M. Kochmann, Inverting the structure–property map of truss metamaterials by deep learning, Proceedings of the National Academy of Sciences 119 (2022) e2111505119.
- [31] L. Zheng, S. Kumar, D. M. Kochmann, Data-driven topology optimization of spinodoid metamaterials with seamlessly tunable anisotropy, Computer Methods in Applied Mechanics and Engineering 383 (2021) 113894.
- [32] M. Mozaffar, R. Bostanabad, W. Chen, K. Ehmann, J. Cao, M. Bessa, Deep learning predicts path-dependent plasticity, Proceedings of the National Academy of Sciences 116 (2019) 26414–26420.
- [33] C. Bonatti, D. Mohr, One for all: Universal material model based on minimal state-space neural networks, Science Advances 7 (2021) eabf3658.
- [34] N. N. Vlassis, R. Ma, W. Sun, Geometric deep learning for computational mechanics part i: Anisotropic hyperelasticity, Computer Methods in Applied Mechanics and Engineering 371 (2020) 113299.
- [35] S. Kumar, D. M. Kochmann, What machine learning can do for computational solid mechanics, in: Current trends and open problems in computational mechanics, Springer, 2022, pp. 275–285.
- [36] P. Thakolkaran, A. Joshi, Y. Zheng, M. Flaschel, L. De Lorenzis, S. Kumar, NN-EUCLID: Deep-learning hyperelasticity without stress data, Journal of the Mechanics and Physics of Solids 169 (2022) 105076.
- [37] R. Lourenço, P. Georgieva, E. Cueto, A. Andrade-Campos, An indirect training approach for implicit constitutive modelling using recurrent neural networks and the virtual fields method, Computer Methods in Applied Mechanics and Engineering 425 (2024) 116961.
- [38] M. Flaschel, S. Kumar, L. De Lorenzis, Unsupervised discovery of interpretable hyperelastic constitutive laws, Computer Methods in Applied Mechanics and Engineering 381 (2021) 113852.

- [39] M. Flaschel, S. Kumar, L. De Lorenzis, Discovering plasticity models without stress data, *npj Computational Materials* 8 (2022) 91.
- [40] E. Marino, M. Flaschel, S. Kumar, L. De Lorenzis, Automated identification of linear viscoelastic constitutive laws with EUCLID, *Mechanics of Materials* 181 (2023) 104643.
- [41] M. Flaschel, S. Kumar, L. De Lorenzis, Automated discovery of generalized standard material models with euclid, *Computer Methods in Applied Mechanics and Engineering* 405 (2023) 115867.
- [42] W. G. Guo, Q. Tian, S. Guo, Y. Guo, A physics-driven deep learning model for process-porosity causal relationship and porosity prediction with interpretability in laser metal deposition, *CIRP Annals* 69 (2020) 205–208.
- [43] C.-T. Chen, G. X. Gu, Physics-Informed Deep-Learning For Elasticity: Forward, Inverse, and Mixed Problems, *Advanced Science* 10 (2023) 2300439. *eprint*: <https://advanced.onlinelibrary.wiley.com/doi/pdf/10.1002/advs.202300439>.
- [44] Q. Nguyen, E. Lejeune, Segmenting mechanically heterogeneous domains via unsupervised learning, *Biomechanics and Modeling in Mechanobiology* 23 (2024) 349–372.
- [45] R. Shi, H. Yang, J. Chen, K. Hackl, S. Avril, Y. He, Deep learning without stress data on the discovery of multi-regional hyperelastic properties, *Computational Mechanics* (2025) 1–30.
- [46] A. Joshi, P. Thakolkaran, Y. Zheng, M. Escande, M. Flaschel, L. De Lorenzis, S. Kumar, Bayesian-euclid: Discovering hyperelastic material laws with uncertainties, *Computer Methods in Applied Mechanics and Engineering* 398 (2022) 115225.
- [47] S. Hartmann, P. Neff, Polyconvexity of generalized polynomial-type hyperelastic strain energy functions for near-incompressibility, *International Journal of Solids and Structures* 40 (2003) 2767–2791.
- [48] A. Isihara, N. Hashitsume, M. Tatibana, Statistical theory of rubber-like elasticity. iv.(two-dimensional stretching), *Journal of Chemical Physics* 19 (1951) 1508–1512.
- [49] L. R. Treloar, The elasticity of a network of long-chain molecules—ii, *Transactions of the Faraday Society* 39 (1943) 241–246.
- [50] V. Biderman, Calculation of rubber parts, *Raschetti na prochnost* 40 (1958).
- [51] D. Haines, W. Wilson, Strain-energy density function for rubberlike materials, *Journal of the Mechanics and Physics of Solids* 27 (1979) 345–360.
- [52] A. G. Baydin, B. A. Pearlmutter, A. A. Radul, J. M. Siskind, Automatic differentiation in machine learning: a survey, *Journal of machine learning research* 18 (2018) 1–43.
- [53] M. Flaschel, S. Kumar, L. De Lorenzis, Unsupervised discovery of interpretable hyperelastic constitutive laws, *Computer Methods in Applied Mechanics and Engineering* 381 (2021) 113852.
- [54] R. Nayek, R. Fuentes, K. Worden, E. J. Cross, On spike-and-slab priors for bayesian equation discovery of nonlinear dynamical systems via sparse linear regression, *Mechanical Systems and Signal Processing* 161 (2021) 107986.
- [55] Q. Nguyen, E. Lejeune, Segmenting mechanically heterogeneous domains via unsupervised learning, *Biomechanics and Modeling in Mechanobiology* 23 (2024) 349–372.
- [56] H. Kobeissi, S. Mohammadzadeh, E. Lejeune, Enhancing mechanical metamodels with a generative model-based augmented training dataset, *Journal of Biomechanical Engineering* 144 (2022) 121002.

- [57] M. Bornert, F. Brémand, P. Doumalin, J.-C. Dupré, M. Fazzini, M. Grédiac, F. Hild, S. Mistou, J. Molimard, J.-J. Orteu, et al., Assessment of digital image correlation measurement errors: methodology and results, *Experimental mechanics* 49 (2009) 353–370.
- [58] C. Saunders, A. Gammerman, V. Vovk, Ridge regression learning algorithm in dual variables, *Proceedings of the 15th International Conference on Machine Learning, ICML* (1998).
- [59] D.-T. Lee, B. J. Schachter, Two algorithms for constructing a delaunay triangulation, *International Journal of Computer & Information Sciences* 9 (1980) 219–242.
- [60] Z. I. Botev, The normal law under linear restrictions: simulation and estimation via minimax tilting, *Journal of the Royal Statistical Society Series B: Statistical Methodology* 79 (2017) 125–148.
- [61] G. Casella, E. I. George, Explaining the gibbs sampler, *The American Statistician* 46 (1992) 167–174.

Supplementary Information – Hetero-EUCLID: Interpretable model discovery for heterogeneous hyperelastic materials using stress-free unsupervised learning

Kanhaiya Lal Chaurasiya^a, Saurav Dutta^a, Siddhant Kumar^b, Akshay Joshi^{a,*}

^a*Department of Mechanical Engineering, Indian Institute of Science, Bengaluru, India*

^b*Department of Materials Science and Engineering, Delft University of Technology, 2628 CD Delft, The Netherlands*

Abstract

We propose a computational framework, Hetero-EUCLID, for segmentation and parameter identification to characterize the full hyperelastic behavior of all constituents of a heterogeneous material. In this work, we leverage the Bayesian-EUCLID (Efficient Unsupervised Constitutive Law Identification and Discovery) framework to efficiently solve the heterogenized formulation through parsimonious model selection using sparsity-promoting priors and Monte Carlo Markov Chain sampling. We utilize experimentally observable 3D surface displacement and boundary-averaged force data generated from Finite Element simulations of non-equibiaxial tension tests on heterogeneous specimens. The framework broadly consists of two steps—residual force-based segmentation, and constitutive parameter identification. We validate and demonstrate the ability of the proposed framework to segment the domain, and characterize the constituent materials on various types of thin square heterogeneous domains. We validate of the framework's ability to segment and characterize materials with various levels of displacement noises and non-native mesh discretizations, i.e., using different meshes for the forward FE simulations and the inverse EUCLID problem. This demonstrates Hetero-EUCLID framework's applicability in Digital Image/Volume Correlation-based experimental scenarios. Furthermore, the proposed framework performs successful segmentation and material characterizations based on data from a single experiment, thereby making it viable for rapid, interpretable model discovery in domains such as aerospace and defense composites and for characterization of selective tissue stiffening in medical conditions such as *fibroatheroma*, *atherosclerosis*, or cancer.

Keywords: Heterogeneous material, data-driven discovery, hyperelastic constitutive models, Bayesian learning, Uncertainty quantification

Contents

S1 Supplementary results	2
S1.1 Case study for moderate noise condition	2
S1.2 Material model parameter prediction using ordinary least squares (OLS) regression	6
S2 Empirical check using strain energy density prediction	7
S2.1 Type 1: Cross-shaped inclusion (High noise)	8
S2.2 Type 2: Split-domain pattern (High noise)	9
S2.3 Type 3: Multiple inner inclusions (High noise)	10
S2.4 Type 4: Cahn-Hilliard pattern (High noise)	11
S2.5 Type 4: Cahn-Hilliard pattern (Non-native mesh)	12

*Corresponding author

Email address: akshayjoshi@iisc.ac.in (Akshay Joshi)

S1. Supplementary results

S1.1. Case study for moderate noise condition

Type 1: Cross-shaped inclusion

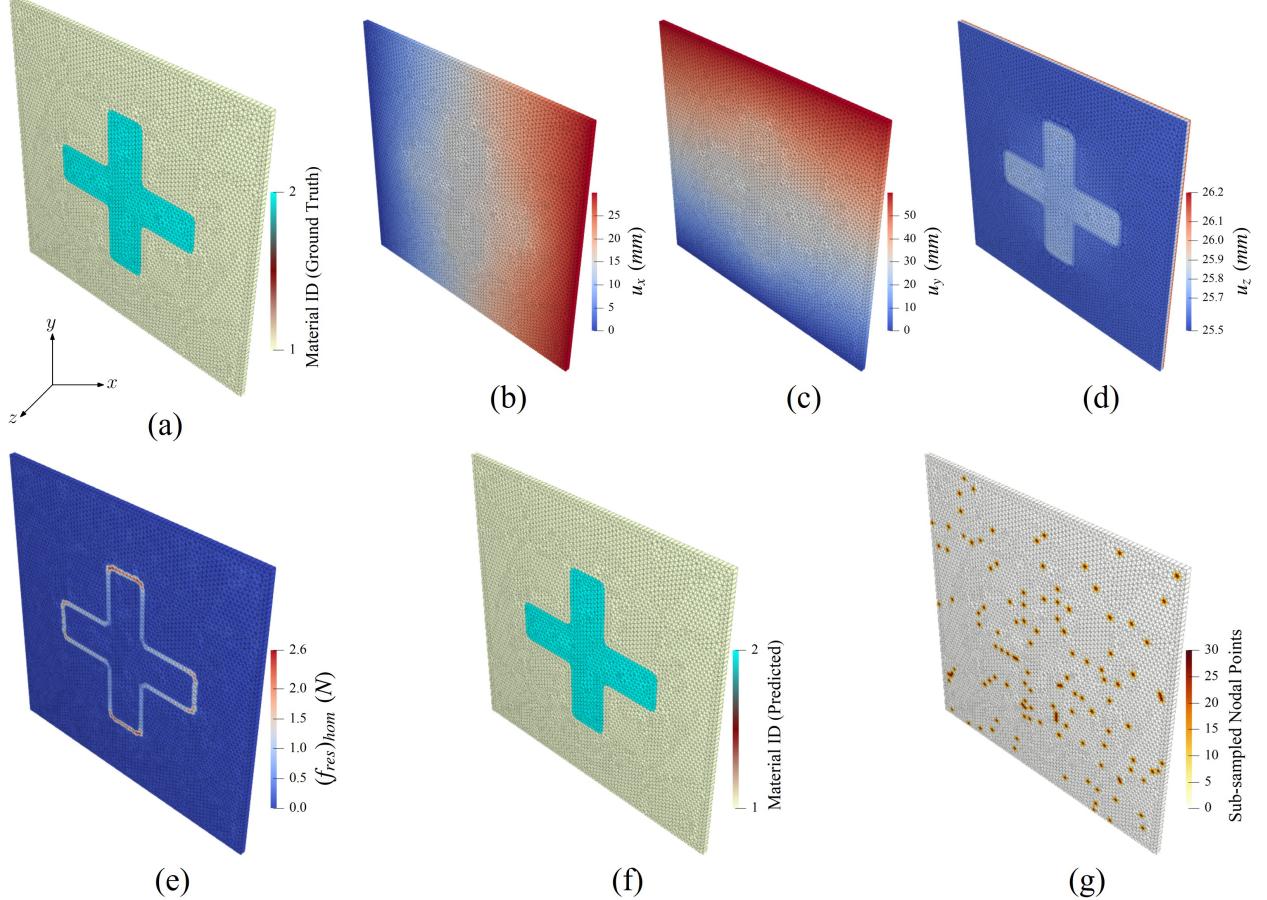


Figure S1: Model discovery for the cross-shaped inclusion pattern (refer to Figure 7a), with sub-figures (b)–(g) corresponding to the moderate noise case. (a) displays the ground truth material segment IDs. (b)–(d) show the displacement fields in the x , y , and z directions, respectively. (e) presents the residual force norm distribution (Equation 24) from the homogenized model solved via ordinary least squares regression. (f) illustrates the predicted material segments from the segmentation algorithm. (g) highlights the sub-sampled nodes from the $\mathcal{D}^{\text{free}}$ subset used to construct the heterogenized model. Table 3 summarizes the ground truth and predicted material parameters under moderate noise conditions obtained by solving the heterogenized model using the Bayesian framework.

Type 2: Split-domain pattern

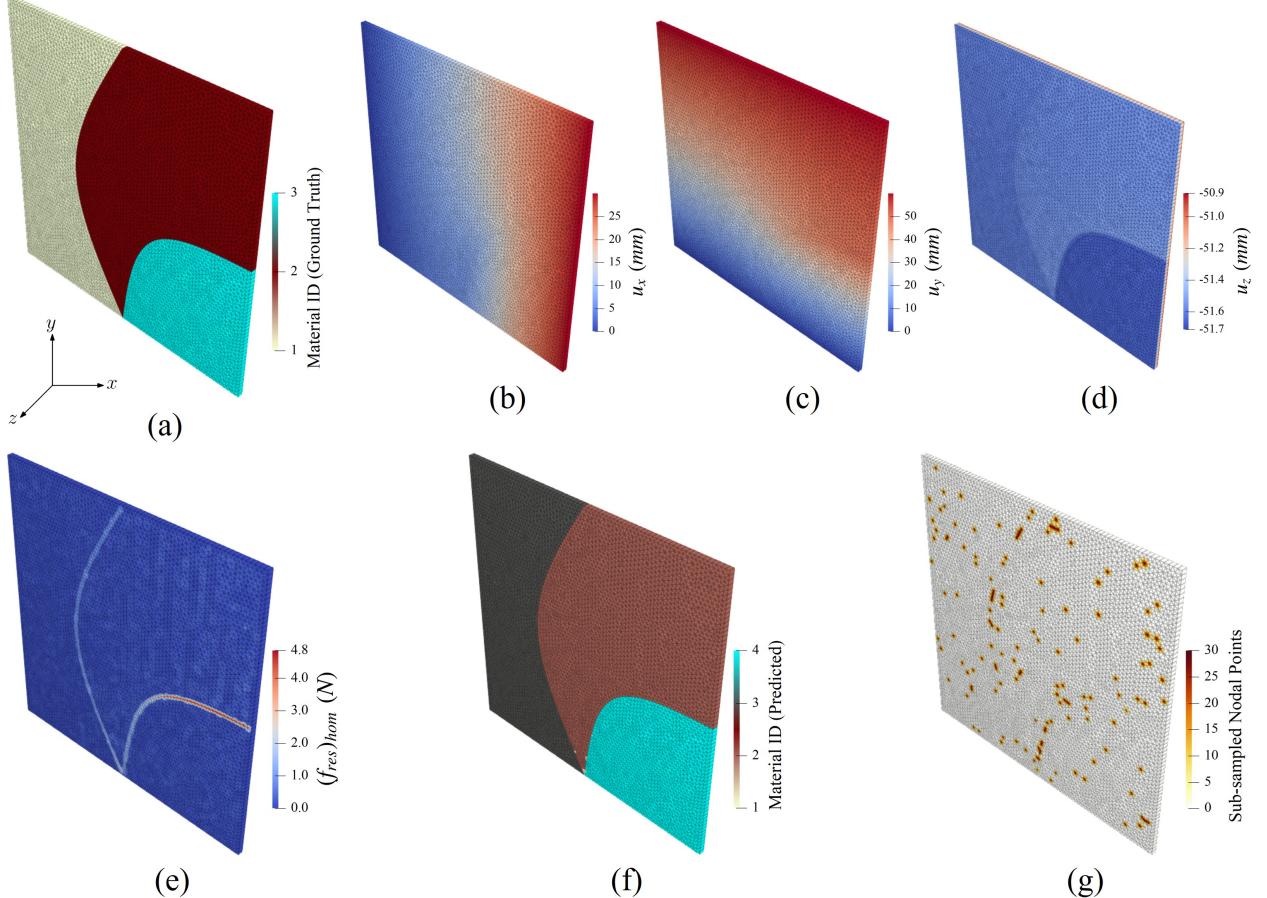


Figure S2: Model discovery for the split-domain pattern (refer to Figure 7b), with sub-figures (b)–(g) corresponding to the moderate noise case. (a) displays the ground truth material segment IDs. (b)–(d) show the displacement fields in the x , y , and z directions, respectively. (e) presents the residual force norm distribution (Equation 24) from the homogenized model solved via ordinary least squares regression. (f) illustrates the predicted material segments from the segmentation algorithm. (g) highlights the sub-sampled nodes from the $\mathcal{D}^{\text{free}}$ subset used to construct the heterogenized model. Table 5 summarizes the ground truth and predicted material parameters under moderate noise conditions obtained by solving the heterogenized model using the Bayesian framework.

Type 3: Multiple inner inclusions

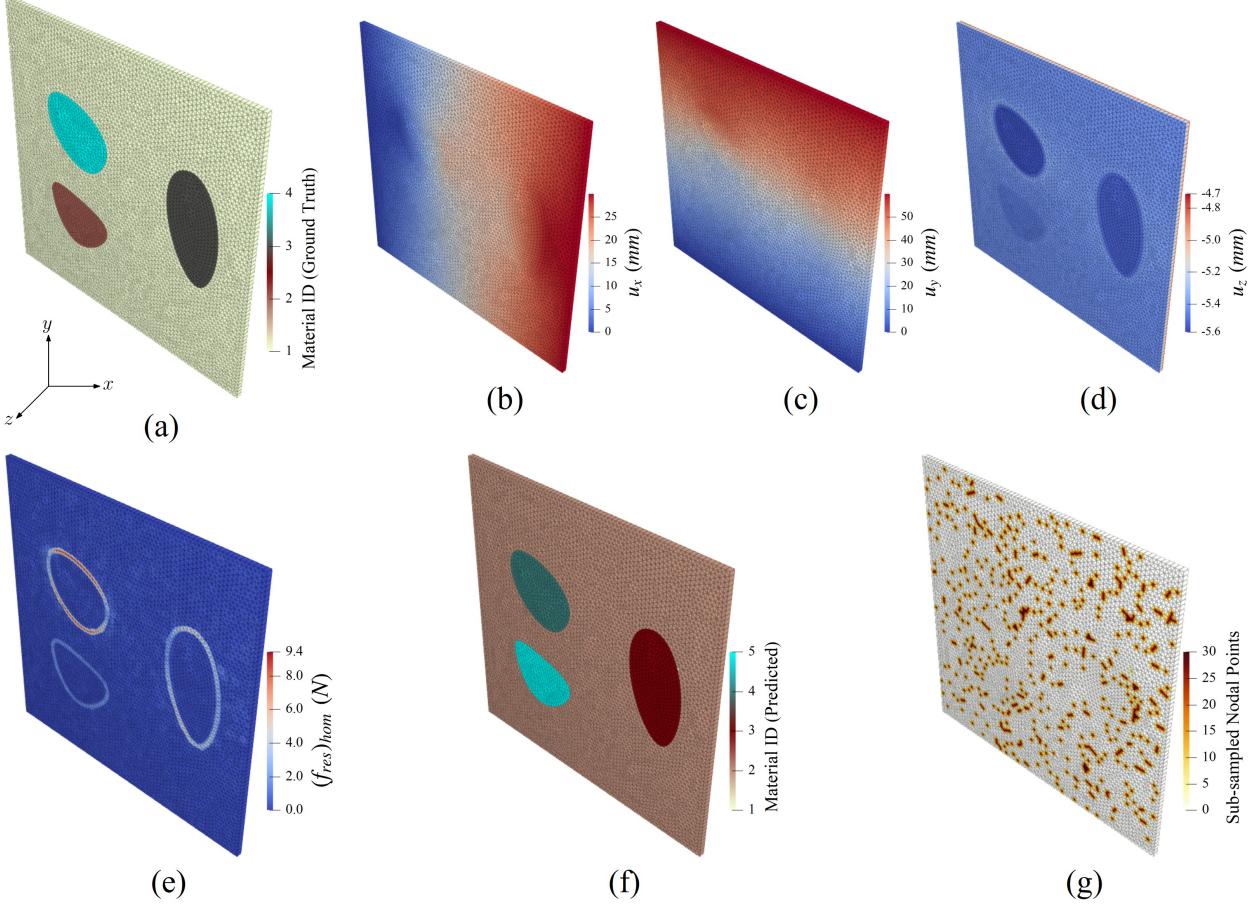


Figure S3: Model discovery for the multiple inner inclusions pattern (refer to Figure 7c), with sub-figures (b)–(g) corresponding to the moderate noise case. (a) displays the ground truth material segment IDs. (b)–(d) show the displacement fields in the x , y , and z directions, respectively. (e) presents the residual force norm distribution (Equation 24) from the homogenized model solved via ordinary least squares regression. (f) illustrates the predicted material segments from the segmentation algorithm. (g) highlights the sub-sampled nodes from the $\mathcal{D}^{\text{free}}$ subset used to construct the heterogenized model. Table 7 summarizes the ground truth and predicted material parameters under moderate noise conditions obtained by solving the heterogenized model using the Bayesian framework.

Type 4: Cahn-Hilliard pattern

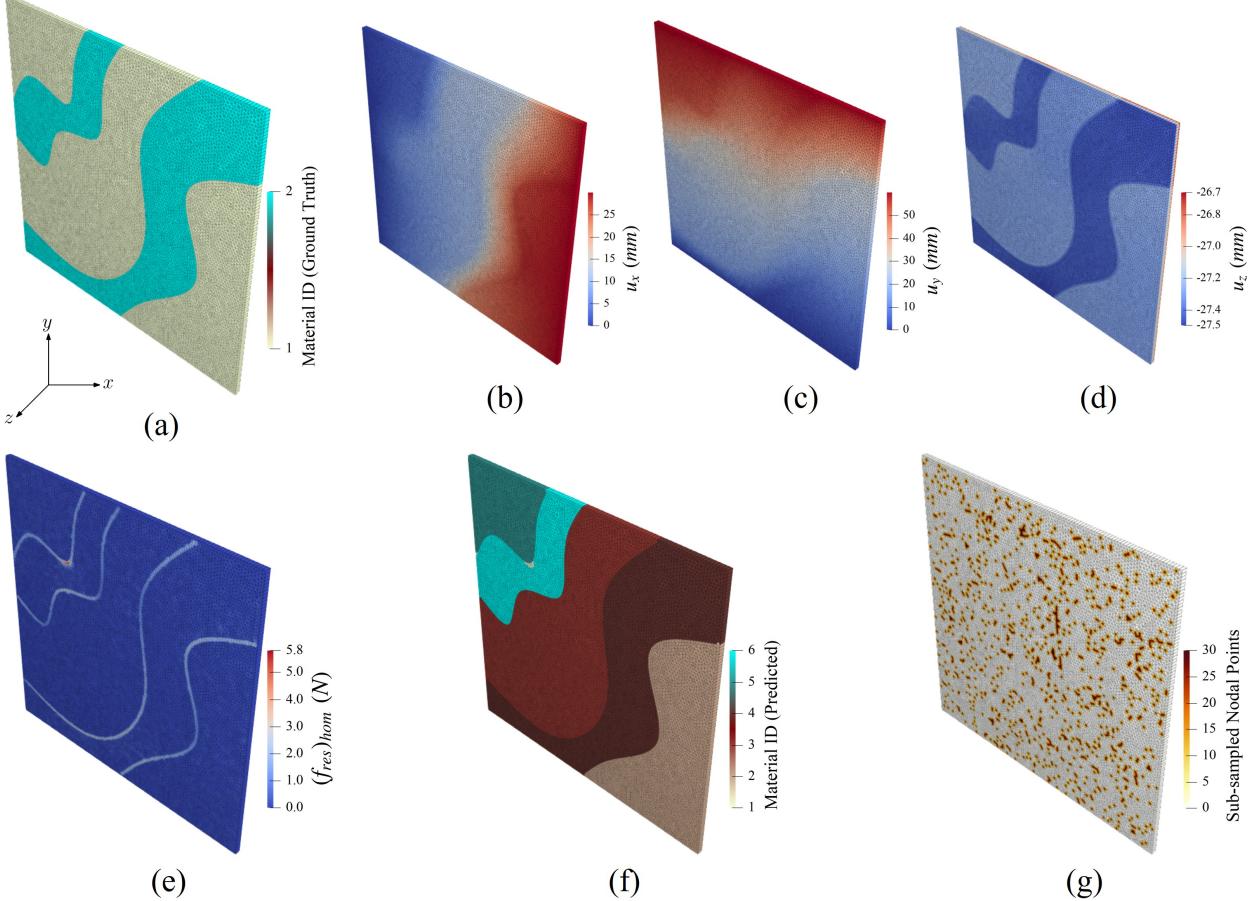


Figure S4: Model discovery for the Cahn-Hilliard pattern (refer to Figure 7d), with sub-figures (b)–(g) corresponding to the moderate noise case. (a) displays the ground truth material segment IDs. (b)–(d) show the displacement fields in the x , y , and z directions, respectively. (e) presents the residual force norm distribution (Equation 24) from the homogenized model solved via ordinary least squares regression. (f) illustrates the predicted material segments from the segmentation algorithm. (g) highlights the sub-sampled nodes from the $\mathcal{D}^{\text{free}}$ subset used to construct the heterogenized model. Table 9 summarizes the ground truth and predicted material parameters under moderate noise conditions obtained by solving the heterogenized model using the Bayesian framework.

S1.2. Material model parameter prediction using ordinary least squares (OLS) regression

Pattern Type: Cross-shaped inclusion Mesh elements (C3D6): 10,442								
Case	Material ID	Constitutive model	θ_1 (MPa)	θ_2 (MPa)	θ_3 (MPa)	θ_4 (MPa)	θ_5 (MPa)	θ_6 (MPa)
Ground Truth	1	NH2 _a	1.80	—	—	—	—	6.00
	2	NH2 _b	5.40	—	—	—	—	15.00
High Noise (using OLS)	2	NH2 _a	1.72	0.04	0.03	-0.01	0.00	5.98
	3	NH2 _b	5.55	-0.10	-0.87	0.29	0.26	14.00

Table S1: True vs. predicted material parameters for cross-shaped inclusion pattern with high noise using ordinary least squares regression

Pattern Type: Split-domain Mesh elements (C3D6): 10,972								
Case	Material ID	Constitutive model	θ_1 (MPa)	θ_2 (MPa)	θ_3 (MPa)	θ_4 (MPa)	θ_5 (MPa)	θ_6 (MPa)
Ground Truth	1	NH2	6.00	—	—	—	—	32.00
	2	ISH	4.00	0.50	0.30	—	—	21.00
High Noise ($\sigma_u = 5 \times 10^{-3}$)	3	HW	1.00	0.15	—	0.02	0.00	10.00
	2	ISH	3.76	0.49	0.95	-0.19	-0.08	32.64
Using ordinary least squares	3	NH2	4.34	0.83	0.41	-0.20	0.00	20.96
	4	HW	-14.48	1.53	2.43	-0.17	-0.11	9.25

Table S2: True vs. predicted material parameters for split-domain pattern with high noise using ordinary least squares regression

Pattern Type: Multiple inner inclusions Mesh elements (C3D6): 10,816								
Case	Material ID	Constitutive model	θ_1 (MPa)	θ_2 (MPa)	θ_3 (MPa)	θ_4 (MPa)	θ_5 (MPa)	θ_6 (MPa)
Ground Truth	1	ISH	4.00	0.50	0.30	—	—	21.00
	2	NH2 _a	5.40	—	—	—	—	15.00
High Noise ($\sigma_u = 5 \times 10^{-3}$)	3	HW	1.00	0.15	—	0.02	0.00	10.00
	4	NH2 _b	1.80	—	—	—	—	6.00
Using ordinary least squares	2	ISH	4.32	0.49	0.17	0.00	0.00	21.05
	5	NH2 _a	-0.75	0.26	0.08	-0.02	0.00	2.04
	3	HW	-23.78	1.19	3.71	-0.09	-0.19	5.78
	4	NH2 _b	-83.63	1.83	27.06	-0.55	-2.63	13.91

Table S3: True vs. predicted material parameters for multiple inner inclusions pattern with high noise using ordinary least squares regression

Pattern Type 4: Cahn-Hilliard Mesh elements (C3D6): 21,040								
Case	Material ID	Constitutive model	θ_1 (MPa)	θ_2 (MPa)	θ_3 (MPa)	θ_4 (MPa)	θ_5 (MPa)	θ_6 (MPa)
Ground Truth	1	NH2 _a	5.40	—	—	—	—	15.00
	2	NH2 _b	1.80	—	—	—	—	6.00
High Noise ($\sigma_u = 5 \times 10^{-3}$) Using ordinary least squares	5	NH2 _a	10.81	-3.40	-3.34	2.86	-0.04	17.75
	6	NH2 _b	1.37	0.03	0.03	0.03	0.00	5.27
	3	NH2 _a	6.81	-0.84	-0.21	0.09	0.01	14.56
	4	NH2 _b	1.56	-0.06	-0.01	0.00	0.00	4.36
	2	NH2 _a	6.05	0.51	0.78	-0.61	-0.04	18.10

Table S4: True vs. predicted material parameters for Cahn-Hilliard pattern with high noise using ordinary least squares regression

S2. Empirical check using strain energy density prediction

We present a series of empirical evaluations conducted to verify the physical admissibility of the discovered material models. For hyperelastic materials, it is well established that the strain energy density must satisfy the condition $W(\mathbf{F}) \geq W(\mathbf{I}) = 0$ for all physically admissible deformation gradients \mathbf{F} . We compute the strain energy density using the discovered models across all heterogeneous specimens shown in Figure 7 under the high noise setting corresponding to Figure 9, 11, 13, 15 and 16. For each constitutive model identified per specimen type, we evaluate both the true and predicted strain energy density along the following six standard deformation paths:

$$\begin{aligned} \mathbf{F}^{\text{UT}}(\gamma) &= \begin{bmatrix} 1 + \gamma & 0 \\ 0 & 1 \end{bmatrix}, & \mathbf{F}^{\text{UC}}(\gamma) &= \begin{bmatrix} 1/(1 + \gamma) & 0 \\ 0 & 1 \end{bmatrix}, & \mathbf{F}^{\text{SS}}(\gamma) &= \begin{bmatrix} 1 & \gamma \\ 0 & 1 \end{bmatrix}, \\ \mathbf{F}^{\text{BT}}(\gamma) &= \begin{bmatrix} 1 + \gamma & 0 \\ 0 & 1 + \gamma \end{bmatrix}, & \mathbf{F}^{\text{BC}}(\gamma) &= \begin{bmatrix} 1/(1 + \gamma) & 0 \\ 0 & 1/(1 + \gamma) \end{bmatrix}, & \mathbf{F}^{\text{PS}}(\gamma) &= \begin{bmatrix} 1 + \gamma & 0 \\ 0 & 1/(1 + \gamma) \end{bmatrix} \end{aligned} \quad (\text{S1})$$

where $\gamma \in [0, 1]$ denotes the deformation path parameter. The abbreviations correspond to different loading modes: uniaxial tensile (UT), uniaxial compressive (UC), biaxial tensile (BT), biaxial compressive (BC), simple shear (SS), and pure shear (PS). For each loading path, we report the predicted strain energy density along with the 95% percentile uncertainty band, and quantify the accuracy using the coefficient of determination R^2 between the predicted and true energy values. These comparisons serve as a comprehensive validation of the physical fidelity of the model across a diverse set of loading conditions.

S2.1. Type 1: Cross-shaped inclusion (High noise)

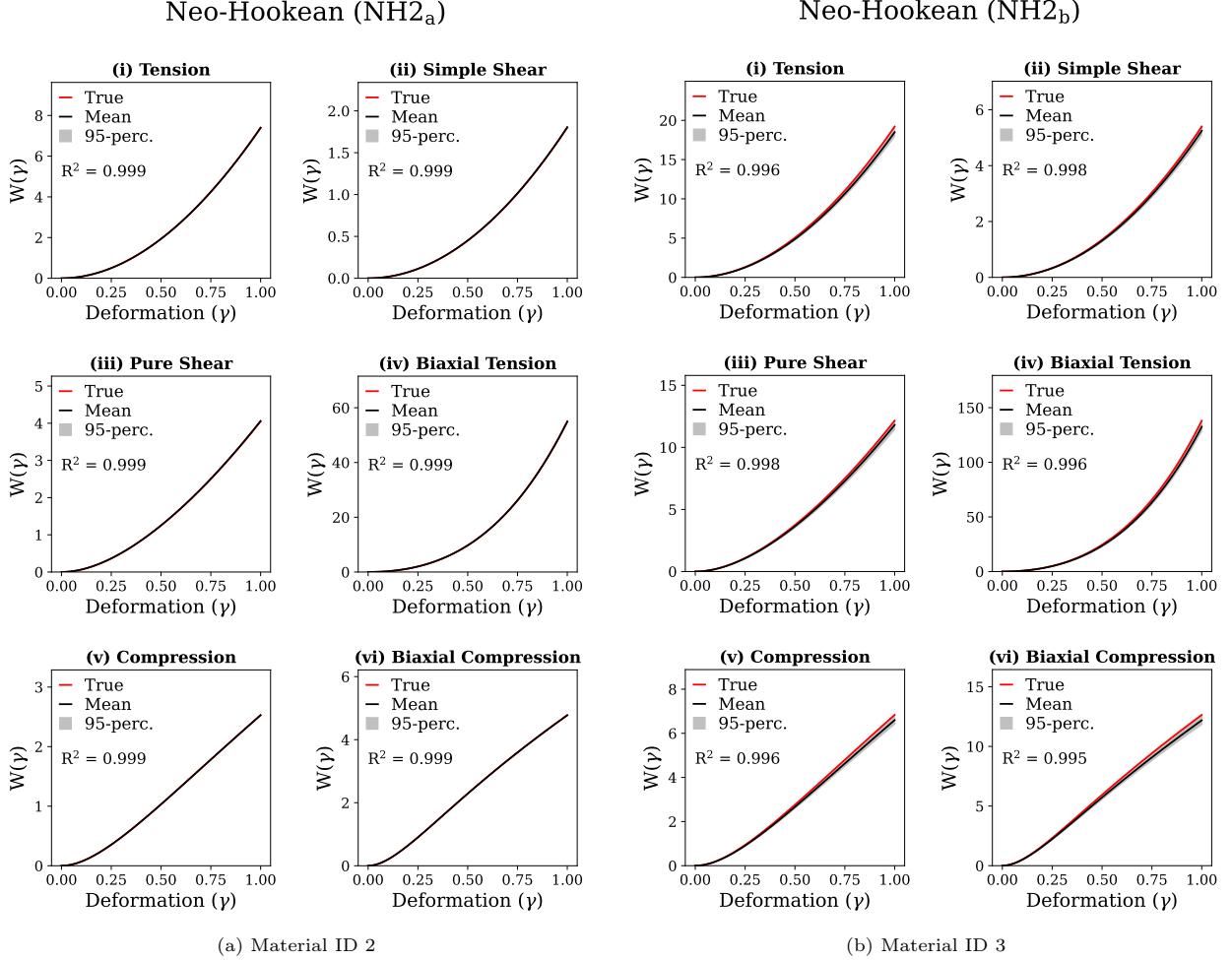


Figure S5: (a) and (b) correspond to the identified material sub-domains for the cross-shaped inclusion pattern under high noise conditions (refer to Figure 9). For each identified material segment ID (a) and (b), subplots (i)–(vi) show strain energy density plots $W(\gamma)$ along various deformation paths (S1), providing a comparison between the discovered constitutive model and the corresponding ground truth.

S2.2. Type 2: Split-domain pattern (High noise)

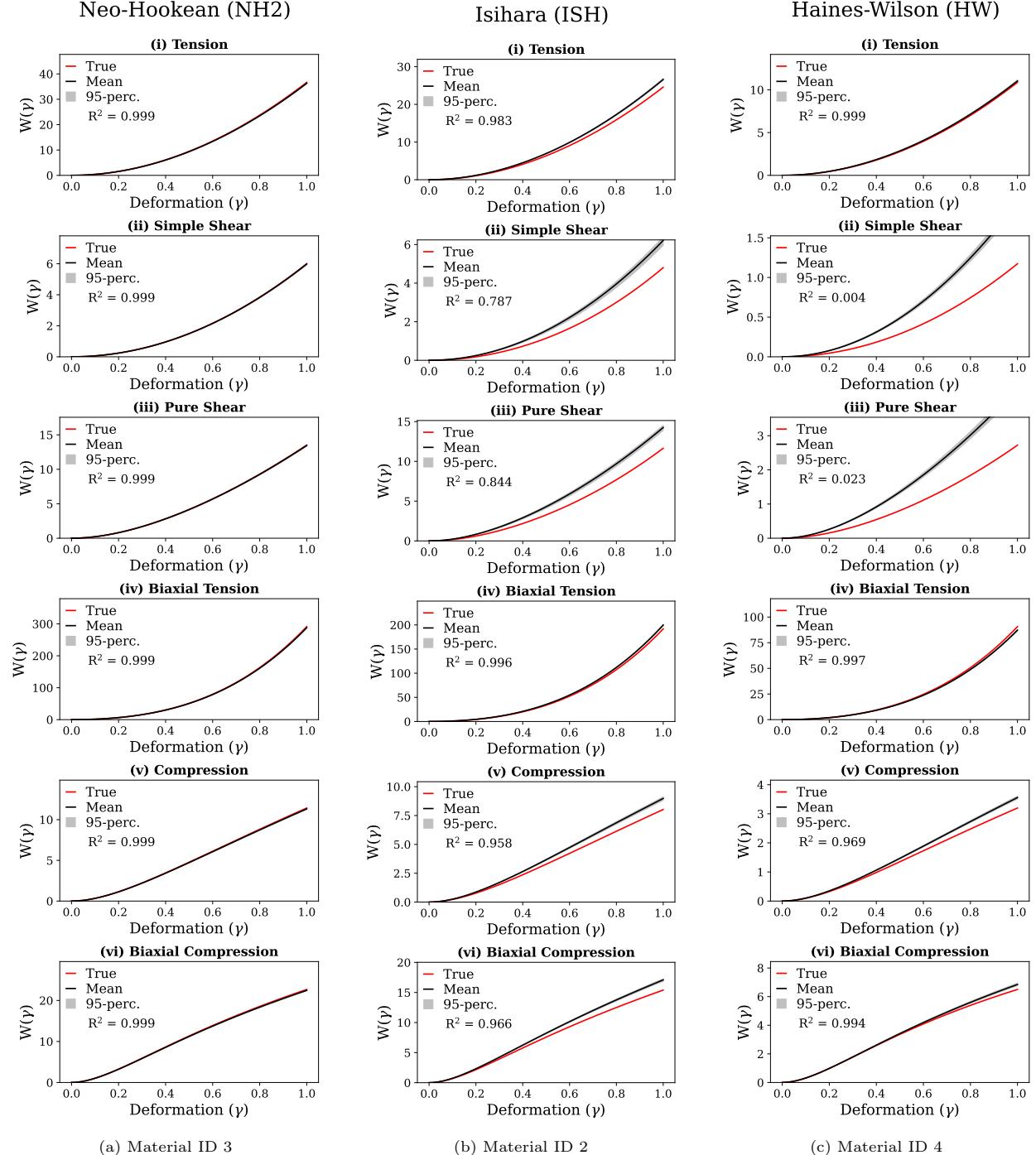


Figure S6: (a)–(c) correspond to the identified material sub-domains for the split-domain pattern under high noise conditions (refer to Figure 11). For each identified material segment ID (a)–(c), subplots (i)–(vi) show strain energy density plots $W(\gamma)$ along various deformation paths (S1), providing a comparison between the discovered constitutive model and the corresponding ground truth.

S2.3. Type 3: Multiple inner inclusions (High noise)

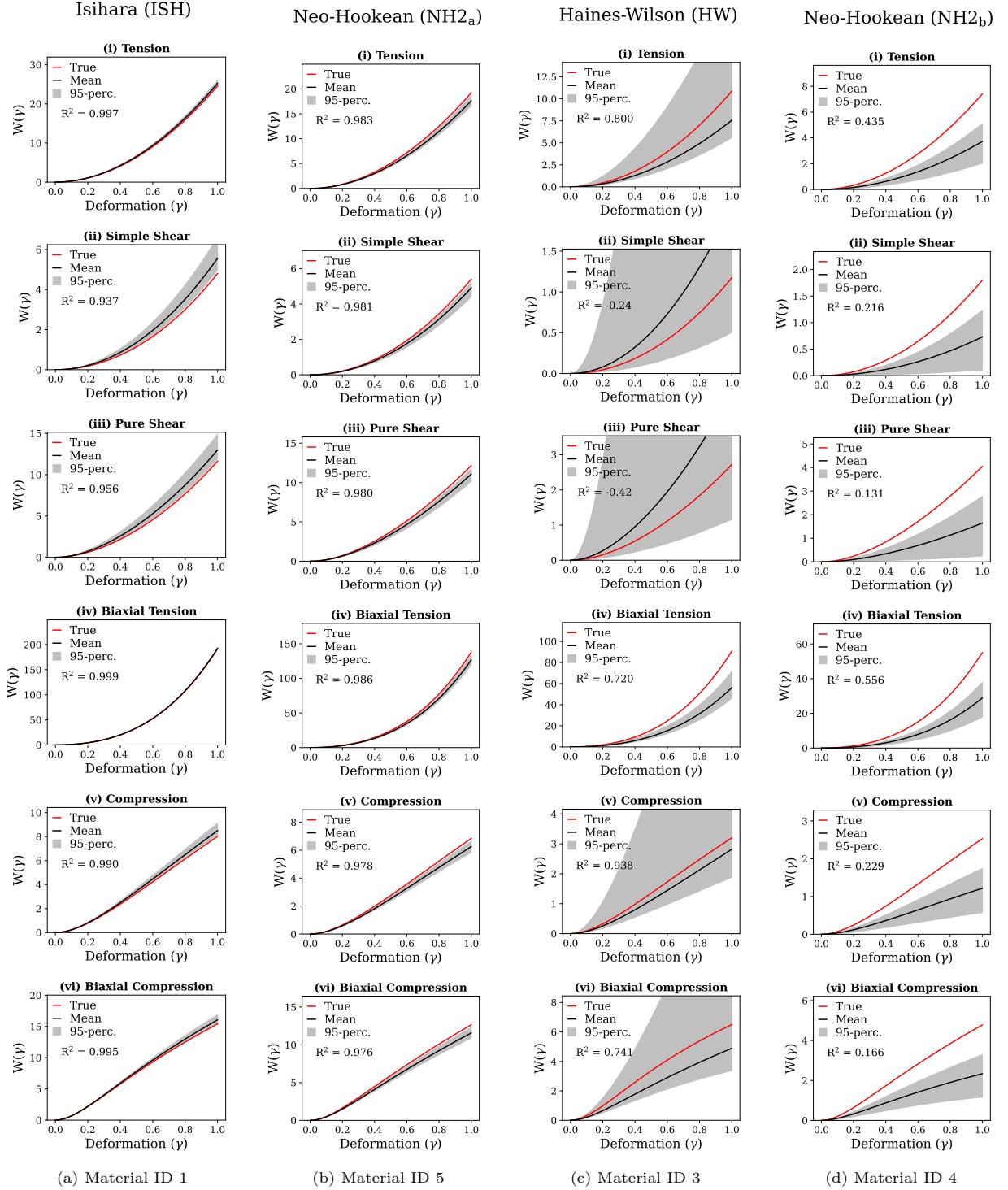


Figure S7: (a)–(d) correspond to the identified material sub-domains for the multiple inner inclusions pattern under high noise conditions (Figure 13). For each identified material segment ID (a)–(d), subplots (i)–(vi) show strain energy density plots $W(\gamma)$ along various deformation paths (S1), providing a comparison between the discovered and true models.

S2.4. Type 4: Cahn-Hilliard pattern (High noise)

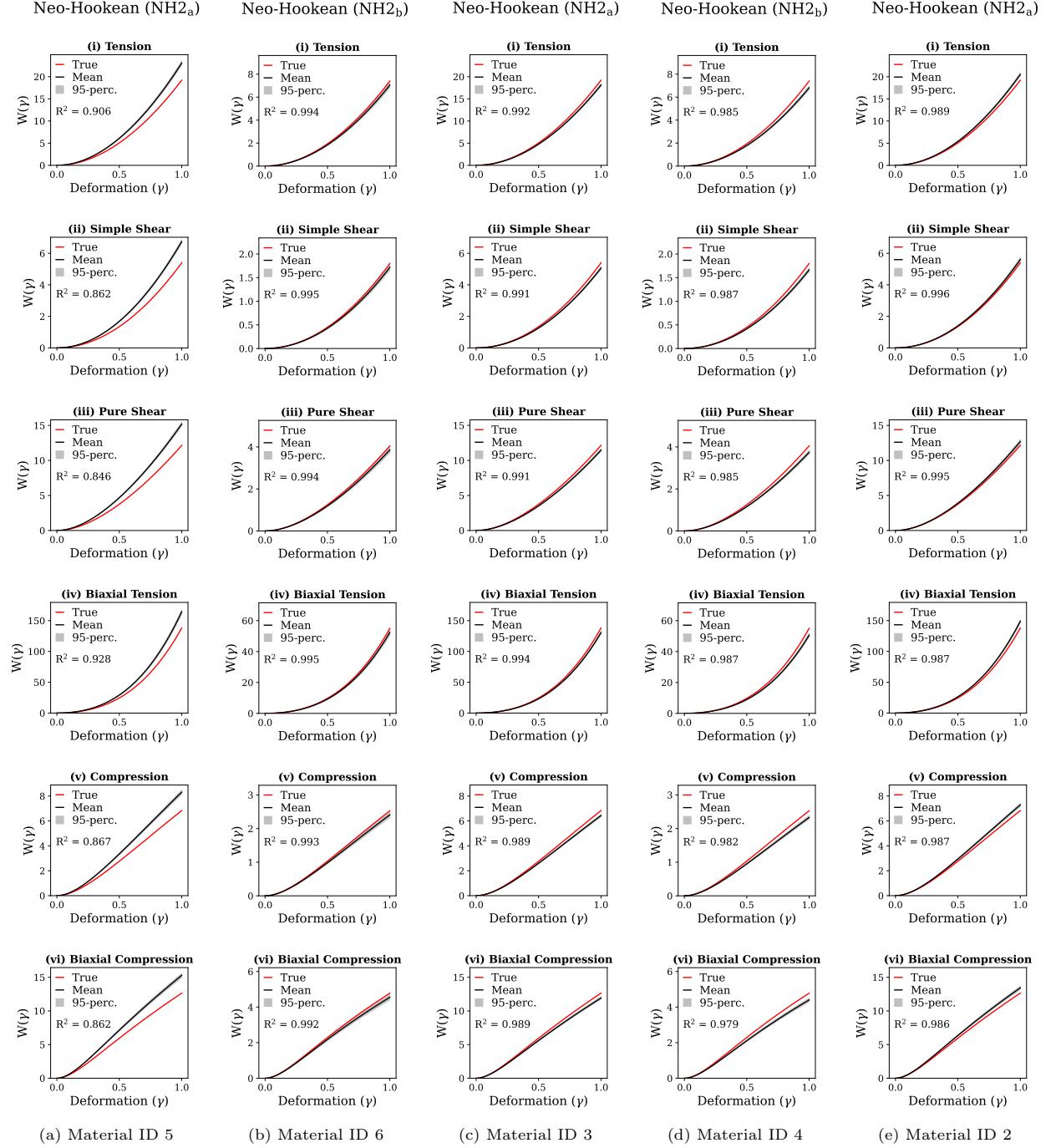


Figure S8: (a)–(e) correspond to the identified material sub-domains for the Cahn-Hilliard pattern under high noise conditions (refer to Figure 15). For each identified material segment ID (a)–(e), subplots (i)–(vi) show strain energy density plots $W(\gamma)$ along various deformation paths (S1), providing a comparison between the discovered constitutive model and the corresponding ground truth.

S2.5. Type 4: Cahn-Hilliard pattern (Non-native mesh)

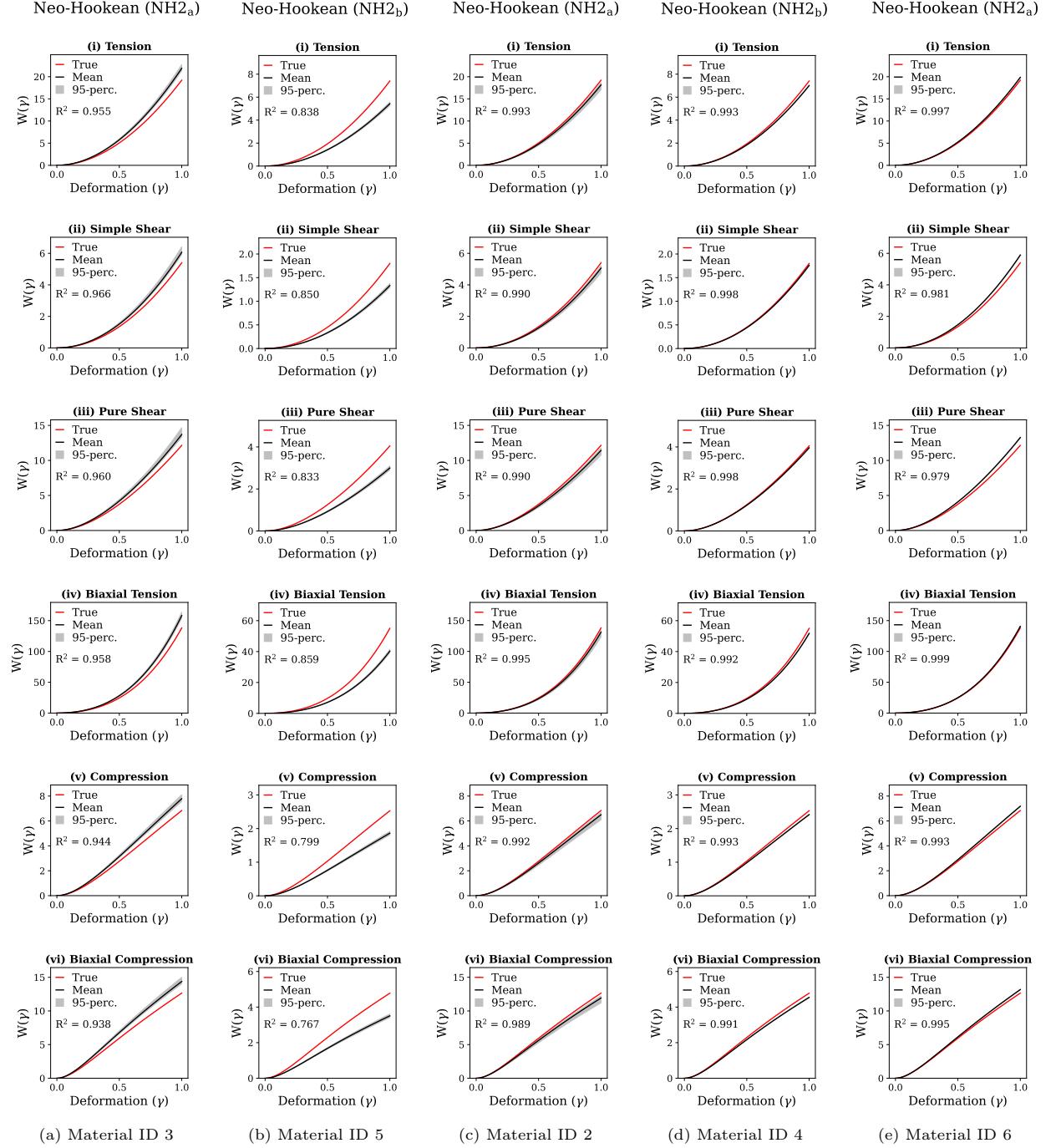


Figure S9: (a)–(e) correspond to the identified material sub-domains for the Cahn-Hilliard pattern with *non-native* meshing (refer to Figure 16). For each identified material segment ID (a)–(e), subplots (i)–(vi) show strain energy density plots $W(\gamma)$ along various deformation paths (S1), providing a comparison between the discovered constitutive model and the corresponding ground truth.