Computer Methods in Applied Mechanics and Engineering 427 (2024) 117061

Contents lists available at ScienceDirect

Comput. Methods Appl. Mech. Engrg.

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

A framework for strategic discovery of credible neural network surrogate models under uncertainty

Pratyush Kumar Singh [a], Kathryn A. Farrell-Maupin [b], Danial Faghihi [a,*]

[a] Department of Mechanical and Aerospace Engineering, University at Buffalo, Buffalo, NY, USA
[b] Sandia National Laboratories, Albuquerque, NM, USA

ARTICLE INFO

Keywords:
Bayesian neural networks
Surrogate modeling
Model validation
Uncertainty quantification
Model plausibility

ABSTRACT

The widespread integration of deep neural networks in developing data-driven surrogate models for high-fidelity simulations of complex physical systems highlights the critical necessity for robust uncertainty quantification techniques and credibility assessment methodologies, ensuring the reliable deployment of surrogate models in consequential decision-making. This study presents the Occam Plausibility Algorithm for surrogate models (OPAL-surrogate), providing a systematic framework to uncover predictive neural network-based surrogate models within the large space of potential models, including various neural network classes and choices of architecture and hyperparameters. The framework is grounded in hierarchical Bayesian inferences and employs model validation tests to evaluate the credibility and prediction reliability of the surrogate models under uncertainty. Leveraging these principles, OPAL-surrogate introduces a systematic and efficient strategy for balancing the trade-off between model complexity, accuracy, and prediction uncertainty. The effectiveness of OPAL-surrogate is demonstrated through two modeling problems, including the deformation of porous materials for building insulation and turbulent combustion flow for ablation of solid fuels within hybrid rocket motors.

1. Introduction

The remarkable advancements in scientific machine learning (SciML) techniques, specifically deep neural networks, ignited an extraordinary revolution in the creation of data-driven surrogate models. These models aim to approximate solutions of high-fidelity physics-based scientific simulations and allowing computational predictions at significantly lower costs. Going beyond the enabler of once-computationally prohibitive outer-loop challenges such as uncertainty quantification, e.g., [1–4], Bayesian inference, e.g., [5,6], design under uncertainty [7,8], digital twins, e.g., [9–13], and optimal experimental design, e.g., [14,15] for complex physical systems, neural network-based surrogate models hold transformative potential in reshaping the formulation and resolution of scientific problems across diverse domains of science, engineering, and medicine.

Despite notable advancements, applying machine learning techniques – originally designed for large data regimes in domains like image processing, computer vision, and natural language processing – encounters significant challenges when directly utilized to construct and establish trust in surrogate models. These obstacles emerge from the inherent spatiotemporal sparsity, limitation, and incompleteness of the scientific data extracted from high-fidelity physical simulations. This intrinsic uncertainty presents substantial challenges to the credibility and prediction reliability of neural network-based surrogate models, as highlighted in works such as [16–19]. The importance of verification, validation, and uncertainty quantification (VVUQ) for physics-based models against

* Corresponding author.
E-mail address: danialfa@buffalo.edu (D. Faghihi).

https://doi.org/10.1016/j.cma.2024.117061
Received 13 March 2024; Received in revised form 11 May 2024; Accepted 11 May 2024
Available online 22 May 2024

experimental measurements is well-established [20–23]. However, as SciML becomes increasingly integrated and deployed into high-consequence decision-making for complex physical systems, there arises a critical need for even more robust UQ techniques and rigorous methodologies to assess the credibility of neural network-based surrogate models. In particular, the first shortcoming lies in the common training approach, based on maximum likelihood parameter estimation, which limits neural networks' robustness against data uncertainty (i.e., adversarial attack), resulting in overfitting and overconfident predictions. The second challenge stems from the existing validation methodologies for neural networks, such as train-and-test approaches that rely on empirical performance assessments with asymptotic guarantees in large data [16,24,25]. Moreover, the interpretability and explainability approaches employed by the machine learning community to build trust in neural network models often resort to heuristic and problem-dependent strategies [16,18,19]. The third and perhaps most pivotal challenge lies in the uncertainty associated with selecting the surrogate model itself. Achieving a delicate balance in the model's complexity is crucial, as overly simplistic models may compromise predictive ability, while excessively complex ones are prone to overfitting the training data, resulting in poor generalization, especially when parameters are estimated via maximum likelihood methods [23,26,27]. Consequently, determining the "best" neural network model becomes challenging in the absence of predefined rules for network architecture and associated hyperparameters. Current trial-and-error architecture selection approaches based on testing data performance are time-consuming and resource-intensive and may not effectively enhance the accuracy and reliability of the surrogate models. Harnessing recent advancements in architecture optimization algorithms and software tools [28–34], when adapted to align the objectives of surrogate modeling, holds the potential to address this challenge effectively.

The Bayesian framework is the cornerstone for successful VVUQ methods, allowing the quantification of uncertainty in model parameters and choice of the model itself in small data regimes, e.g., [6,35–42] and serving as a robust foundation for validating model predictions, e.g., [20–23,43,44]. The seminal contributions of Mackay et al. [26,45] laid the groundwork for the adoption of the Bayesian framework in inferring neural network parameters and hyperparameters, inspiring subsequent developments [27,46]. Despite these early contributions, Bayesian neural networks (BayesNN) only recently gained recognition in the SciML community, e.g., [2,47–53]. The delayed adoption can be attributed to the incomplete understanding of UQ methods for neural networks as well as the computational complexities associated with high-dimensional parameter spaces in these models. BayesNN provides remarkable advantages, including alleviating overfitting, preventing overconfident parameter estimation in small and uncertain datasets, and the ability to quantify prediction uncertainty. This study emphasizes an additional benefit of BayesNN in surrogate modeling by relaxing rigid constraints on model complexity. This flexibility enables the retention of a sufficient number of parameters, which is crucial for capturing the underlying multiscale structure inherent in physics-based simulations. Despite these merits, BayesNN surrogate modeling faces formidable challenges, particularly in (i) selecting a specific network architecture and hyperparameters, as it is challenging to assert that the chosen models align with prior beliefs about the problem, and (ii) developing methodologies to rigorously assess the credibility and prediction reliability of the surrogate models under uncertainty.

This contribution introduces the Occam Plausibility Algorithm for surrogate models (OPAL-surrogate), a systematic framework designed to discover predictive neural network-based surrogate models for high-fidelity physical simulations. The name is inspired by the principle of Occam's Razor, advocating the preference for simpler models over unnecessary complex ones, guided by the notion of model plausibility as a basis for its effectiveness in explaining the given data. OPAL-surrogate is grounded in hierarchical Bayesian inferences, enabling a systematic determination of the probability distributions of the network parameters and hyperparameters and measures for comparing various neural network models. Moreover, it adheres to the principles of Bayesian model validation to assess the credibility and prediction reliability of the surrogate model. Leveraging these methodologies, OPAL-surrogate presents a strategy to adaptively adjust model complexity, utilizing a combination of bottom-up and top-down approaches until predefined validation criteria are met. Consequently, within the wide space of potential BayesNN models involving choices of architecture and hyperparameters, OPAL-surrogate identifies the "best" predictive surrogate model by balancing the trade-off between model complexity and prediction uncertainty. The effectiveness of OPAL-surrogate is demonstrated via two modeling problems in solid mechanics and computational fluid dynamics, identifying credible surrogate models for reliable prediction of quantities of interest (QoIs).

Following this introduction, Section 2 provides a comprehensive overview of Bayesian learning for neural networks along with an efficient and scalable solution algorithm. Section 3 delves into the definition of the BayesNN model and hierarchical inference of various classes of neural network parameters. The steps involved in the OPAL-surrogate for discovering predictive BayesNN surrogate models are outlined in Section 4. Section 5 shows numerical examples, including the identification of BayesNN surrogate models for the elastic deformation of porous materials with random microstructures and the direct numerical simulation of turbulent combustion flow for shear-induced ablation of solid fuels within hybrid rocket motors. Concluding remarks can be found in Section 6.

2. Neural networks learning as probabilistic inference

A neural network is a nonlinear map $\boldsymbol{g} : \mathbb{R}^{d_i} \to \mathbb{R}^{d_o}$ from the inputs $\boldsymbol{x} \in \mathbb{R}^{d_i}$ to the outputs $\boldsymbol{u}_{\boldsymbol{\theta}} \in \mathbb{R}^{d_o}$, representing a continuously parameterized function. The functional form of a feed-forward neural network with D number of layers is expressed as follows,

$$\boldsymbol{u}_{\boldsymbol{\theta}}(\boldsymbol{x}) = f^{(D)}(\boldsymbol{w}^{(D)} \cdots f^{(2)}(\boldsymbol{w}^{(2)} f^{(1)}(\boldsymbol{w}^{(1)}\boldsymbol{x} + \boldsymbol{b}^{(1)}) + \boldsymbol{b}^{(2)}) \cdots + \boldsymbol{b}^{(D)}). \tag{1}$$

Here, $\boldsymbol{w}^{(\ell)}$ represents the weight vector connecting layer $\ell - 1$ to layer ℓ, $\boldsymbol{b}^{(\ell)}$ denotes the biases in layer ℓ, and $f^{(\ell)}$ denotes activation functions applied element-wise. The set of weights and biases collectively forms the network parameters $\boldsymbol{\theta} \in \mathbb{R}^P$, and

common activation functions include Hyperbolic Tangent (*Tanh*), Rectified Linear Unit (*ReLU*), Leaky Rectified Linear Unit (*Leaky ReLU*), and Logistic (*Sigmoid*) functions. The output in each layer ℓ can be represented as

$$\boldsymbol{a}^{(\ell)} = \boldsymbol{w}^{(\ell)}\boldsymbol{z}^{(\ell-1)} + \boldsymbol{b}^{(\ell)}; \quad \boldsymbol{z}^{(\ell)} = f^{(\ell)}(\boldsymbol{a}^{(\ell)}); \quad 1 \leq \ell < D, \tag{2}$$

where $\boldsymbol{a}^{(\ell)}$ is the pre-activation and $\boldsymbol{z}^{(\ell)}$ denotes the activation values. Standard training based on the maximum likelihood estimate often leads to overconfidence in parameter values and ignoring inherent uncertainty in model predictions. However, BayesNN accounts for this uncertainty by considering probability distribution functions (PDFs) over the parameters, inferred from data $\boldsymbol{D} = \{\boldsymbol{x}_{\boldsymbol{D}}^j, \boldsymbol{u}_{\boldsymbol{D}}^j\}_{j=1}^{N_D}$ using Bayes' theorem,

$$\pi_{post}(\boldsymbol{\theta}|\boldsymbol{D}) = \frac{\pi_{like}(\boldsymbol{D}|\boldsymbol{\theta})\pi_{pr}(\boldsymbol{\theta})}{\pi_{evid}(\boldsymbol{D})}. \tag{3}$$

Here, $\pi_{post}(\boldsymbol{\theta}|\boldsymbol{D})$ represents the posterior PDF, updating the prior PDF $\pi_{pr}(\boldsymbol{\theta})$ given observational data and the evidence PDF $\pi_{evid}(\boldsymbol{D})$ serves as the normalization factor. Additionally, the likelihood PDF $\pi_{like}(\boldsymbol{D}|\boldsymbol{\theta})$ is derived from a noise model depicting the discrepancy between the data and neural network output. Adhering to the maximum entropy principle with constraints on the mean and variance of parameters, we adopt a Gaussian prior $\pi_{pr}(\boldsymbol{\theta}) = \mathcal{N}(\bar{\boldsymbol{\theta}}, \boldsymbol{\Gamma}_{pr})$. Unless derived from pre-training or historical datasets, and without loss of generality, we assume $\bar{\boldsymbol{\theta}} = \boldsymbol{0}$ and prior covariance is $\boldsymbol{\Gamma}_{pr} = (\sigma^{\text{pr}})^2\mathbf{I}$ where σ^{pr} is the prior hyper-parameter. Additionally, we consider an additive noise model $\boldsymbol{u}_D = \boldsymbol{u}_{\boldsymbol{\theta}}(\boldsymbol{x}_{\boldsymbol{D}}) + \boldsymbol{\eta}$ where $\boldsymbol{\eta}$ is the total error, including both data uncertainty and network inadequacy in representing the data. Assuming a zero-mean Gaussian distribution for the total error, $\boldsymbol{\eta} \sim \mathcal{N}\left(\boldsymbol{0}, (\sigma^{\text{noise}})^2\mathbf{I}\right)$ and considering independent and identically distributed (iid) data, the likelihood PDF is expressed as,

$$\pi_{\text{like}}(\boldsymbol{D}|\boldsymbol{\theta}) = \prod_{j=1}^{N_D} \frac{1}{\sqrt{2\pi}\sigma^{\text{noise}}} \exp\left[-\frac{(\boldsymbol{u}_{\boldsymbol{D}}^j - \boldsymbol{u}_{\boldsymbol{\theta}}(\boldsymbol{x}_{\boldsymbol{D}}^j))^2}{2(\sigma^{\text{noise}})^2}\right], \tag{4}$$

where σ^{noise} is called noise hyper-parameter. A point estimate of the most probable value for the parameter, considering both the observed data and the prior PDFs, is referred to as the Maximum A Posteriori (MAP) estimate such that,

$$\boldsymbol{\theta}_{MAP} = \underset{\boldsymbol{\theta}\in\Theta}{\text{argmax}}\ \pi_{post}(\boldsymbol{\theta}|\boldsymbol{D}). \tag{5}$$

2.1. Bayesian solution: Laplace approximation to the posterior

Bayesian inference for neural networks poses substantial computational challenges, primarily due to the high-dimensional parameter space (number of weights) and the potential complexity of the posterior distribution geometry. To illustrate the proposed framework for BayesNN surrogate modeling in this work, we leverage an efficient Bayesian solution based on Laplace approximation (LA) to the posterior, offering scalability with respect to the parameter dimensions. The utilization of LA within BayesNN draws from seminal works of [45,46,54], and recently, its computational efficiency has been improved to accommodate larger networks [55–59]. This approach involves approximating the posterior distribution by linearizing it around its dominant mode (the MAP estimate in (5)), followed by the application of Laplace's method to evaluate the resulting normalization factor, yielding,

$$\pi_{post}(\boldsymbol{\theta}|\boldsymbol{D}) \approx \frac{1}{\sqrt{(2\pi)^P \det(\boldsymbol{H})}} \exp\left[-\frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta}_{MAP})^T \boldsymbol{H}(\boldsymbol{\theta} - \boldsymbol{\theta}_{MAP})\right]. \tag{6}$$

Here, $\boldsymbol{H} = -\nabla^2_{\boldsymbol{\theta}\boldsymbol{\theta}} \ln \pi_{post}(\boldsymbol{\theta}|\boldsymbol{D})|_{\boldsymbol{\theta}=\boldsymbol{\theta}_{MAP}}$ is the positive (semi-)definite Hessian of the negative log-posterior, describing its local curvature at the MAP point, and P is the number of parameters $\boldsymbol{\theta}$. The relation (6) demonstrates that the posterior PDF in approximated by a multivariate normal distribution $\mathcal{N}(\boldsymbol{\theta}_{MAP}, \boldsymbol{\Gamma}_{post})$. The mean of this distribution is the MAP point $\boldsymbol{\theta}_{MAP}$, and the posterior covariance matrix $\boldsymbol{\Gamma}_{post} = \boldsymbol{H}^{-1}$ is given by the inverse of the Hessian matrix evaluated at $\boldsymbol{\theta}_{MAP}$. The MAP estimate for the parameters can be computed by minimizing the negative log-posterior, which can be expressed equivalently through the likelihood and prior PDFs as,

$$\boldsymbol{\theta}_{MAP} = \underset{\boldsymbol{\theta}}{\text{argmin}}\ \{-\ln \pi_{post}(\boldsymbol{\theta}|\boldsymbol{D})\} = \underset{\boldsymbol{\theta}}{\text{argmin}}\ \{-\ln \pi_{like}(\boldsymbol{D}|\boldsymbol{\theta}) - \ln \pi_{pr}(\boldsymbol{\theta})\}. \tag{7}$$

The solution to the above optimization problem is analogous to deterministic neural network training with an updated loss function. Furthermore, the posterior covariance can be expressed as,

$$\boldsymbol{\Gamma}_{post} = \boldsymbol{H}^{-1} = \left(\boldsymbol{H}_{lnlike} + \boldsymbol{\Gamma}_{pr}^{-1}\right)^{-1}, \tag{8}$$

where $\boldsymbol{\Gamma}_{pr}$ is the prior covariances, and $\boldsymbol{H}_{lnlike} = -\nabla^2_{\boldsymbol{\theta}\boldsymbol{\theta}} \ln \pi_{like}(\boldsymbol{D}|\boldsymbol{\theta})\big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_{MAP}}$ denotes the Hessian of log-likelihood evaluated at MAP point. To this end, the efficient computation of the log-likelihood Hessian is crucial for the calculation of the posterior covariance, as the prior terms are typically straightforward.

2.2. Scalable algorithm via Kronecker-factored Laplace approximation

The computation of the posterior covariance encounters a significant challenge due to the large size of the $\boldsymbol{H}_{lnlike} \in \mathbb{R}^{P\times P}$ matrix resulting from the high-dimensional parameter space in neural networks. This matrix is often non-diagonal, not guaranteed to be positive semi-definite, and may lack sparsity. To develop a scalable solution algorithm for Bayesian inference, we adopt the

Kronecker factored representation of the Hessian proposed by [60,61]. This approach leverages the structure of neural network parameters, enabling explicit calculation, inversion, and storage of this matrix in a tractable manner.

Due to the intractability of the $\boldsymbol{H}_{lnlike}$, it is common practice to resort to its generalized Gauss–Newton approximation. The resulting symmetric positive and semi-definite approximated Hessian matrix is given by,

$$\boldsymbol{H}_{lnlike} \approx \boldsymbol{J}^T \boldsymbol{L} \boldsymbol{J}, \tag{9}$$

where the matrix $\boldsymbol{J} \in \mathbb{R}^{N_D d_o \times P}$ is formed by concatenating N_D Jacobian matrices of the network with respect to the parameters $\boldsymbol{j}(\boldsymbol{x}_{\boldsymbol{D}}^j) = \nabla_{\boldsymbol{\theta}} \boldsymbol{u}_{\boldsymbol{\theta}}(\boldsymbol{x}_{\boldsymbol{D}}^j)|_{\boldsymbol{\theta}=\boldsymbol{\theta}_{MAP}} \in \mathbb{R}^{d_o \times P}$ for all $\{\boldsymbol{x}_{\boldsymbol{D}}^j\}_{j=1}^{N_D}$ in the data set, and $\boldsymbol{L} \in \mathbb{R}^{N_D d_o \times N_D d_o}$ is a block-diagonal matrix with blocks $\boldsymbol{l}(\boldsymbol{u}_{\boldsymbol{D}}^j, \boldsymbol{u}_{\boldsymbol{\theta}}(\boldsymbol{x}_{\boldsymbol{D}}^j)) = -\nabla_{\boldsymbol{gg}}^2 \ln \pi_{like}(D^j|\boldsymbol{\theta})|_{\boldsymbol{\theta}=\boldsymbol{\theta}_{MAP}} \in \mathbb{R}^{d_o \times d_o}$, where the likelihood of the jth data is defined as $\pi_{\text{like}}(D^j|\boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi}\sigma^{\text{noise}}} \exp\left[-\frac{(\boldsymbol{u}_{\boldsymbol{D}}^j - \boldsymbol{u}_{\boldsymbol{\theta}}(\boldsymbol{x}_{\boldsymbol{D}}^j))^2}{2(\sigma^{\text{noise}})^2}\right]$. Despite the approximation in (9), the log-likelihood Hessian remains a full matrix that needs to be inverted. To address the associated computational challenge, we employ a block-diagonal approximation in which the $\boldsymbol{H}_{lnlike}$ matrix is divided into blocks, each corresponding to all weights of one layer of the network. This approach neglects covariance between layers but retains covariance within each layer. We further leverage the Kronecker factored representation of the layer-wise Hessian, denoted as $\boldsymbol{H}_{lnlike}^{(\ell)}$. Proposed by [60,61], the Kronecker-factored representation relies on approximating the sum of Kronecker products over all data points by a Kronecker product of sums. Consequently, the log-likelihood Hessian for the ℓth layer of the neural network can be represented as,

$$[\boldsymbol{J}^T \boldsymbol{L} \boldsymbol{J}]^{(\ell)} \approx \boldsymbol{Q}^{(\ell)} \otimes \boldsymbol{R}^{(\ell)}, \tag{10}$$

where $\boldsymbol{Q}^{(\ell)} = \boldsymbol{z}^{(\ell-1)}(\boldsymbol{z}^{(\ell-1)})^T$ is the uncentered covariance of the activations and $\boldsymbol{R}^{(\ell)} = -\nabla^2_{\boldsymbol{a}^{(\ell)}\boldsymbol{a}^{(\ell)}} \ln \pi_{like}(\boldsymbol{D}|\boldsymbol{\theta})$ and $\boldsymbol{z}^{(\ell)}$ and $\boldsymbol{a}^{(\ell)}$ are the activation values and the pre-activation for layer ℓ defined in (2). As detailed in [61], computing $\boldsymbol{R}^{(\ell)}$ involves a recursive procedure that starts at the output layer, followed by backward propagation through the network in a single pass for each data point, and then averaging across all data points. The matrices $\boldsymbol{Q}^{(\ell)} \in \mathbb{R}^{d_i^{(\ell)} \times d_i^{(\ell)}}$ and $\boldsymbol{R}^{(\ell)} \in \mathbb{R}^{d_o^{(\ell)} \times d_o^{(\ell)}}$ depends in the dimensionality of the ℓth layer's input $d_i^{(\ell)}$ and output $d_o^{(\ell)}$, and both are positive semidefinite. Consequently, the Kronecker-factored approximation offers a significant computational advantage by decomposing the layer-wise log-likelihood Hessian into smaller matrices [55,58]. By substituting (9) and (10) into (8), the posterior covariance at the ℓth layer is approximated as,

$$\boldsymbol{\Gamma}_{post}^{(\ell)} \approx \left(\boldsymbol{Q}^{(\ell)} \otimes \boldsymbol{R}^{(\ell)} + \boldsymbol{\Gamma}_{pr}^{(\ell)^{-1}}\right)^{-1}, \tag{11}$$

while the entire term under inversion does not necessarily allow a Kronecker-factored representation. To preserve a Kronecker factored structure, [55] suggested approximating the posterior covariance by introducing the effect of the prior as damping factors to $\boldsymbol{Q}^{(\ell)}$ and $\boldsymbol{R}^{(\ell)}$. While such dampening is common in optimization methods employing Kronecker-factored Hessian approximations [60], applying it to a posterior approximation artificially increases the posterior concentration. To prevent the introduction of additional approximations that might overshadow the impact of the prior PDF on inference, we utilize the eigendecompositions $\boldsymbol{Q}^{(\ell)} = \boldsymbol{M}_{Q^{(\ell)}} \boldsymbol{\Lambda}_{Q^{(\ell)}} \boldsymbol{M}_{Q^{(\ell)}}^T$ and $\boldsymbol{R}^{(\ell)} = \boldsymbol{M}_{R^{(\ell)}} \boldsymbol{\Lambda}_{R^{(\ell)}} \boldsymbol{M}_{R^{(\ell)}}^T$. Here, $\boldsymbol{\Lambda}_{Q^{(\ell)}} = \text{diag}(\boldsymbol{q}^{(\ell)})$ and $\boldsymbol{\Lambda}_{R^{(\ell)}} = \text{diag}(\boldsymbol{r}^{(\ell)})$ where $\boldsymbol{q}^{(\ell)}$ and $\boldsymbol{r}^{(\ell)}$ are the eigenvalues of $\boldsymbol{Q}^{(\ell)}$ and $\boldsymbol{R}^{(\ell)}$, respectively. Substituting the eigendecompositions into (11) and considering the ℓth block-diagonal entry of prior covariance as $\boldsymbol{\Gamma}_{pr}^{(\ell)} = (\sigma^{\text{pr}(\ell)})^2 \boldsymbol{I}^{(\ell)}$, results in [57],

$$\boldsymbol{\Gamma}_{post}^{(\ell)} \approx \left((\boldsymbol{M}_{Q^{(\ell)}} \otimes \boldsymbol{M}_{R^{(\ell)}})(\boldsymbol{\Lambda}_{Q^{(\ell)}} \otimes \boldsymbol{\Lambda}_{R^{(\ell)}} + (\sigma^{\text{pr}(\ell)})^{-2} \boldsymbol{I}^{(\ell)})(\boldsymbol{M}_{Q^{(\ell)}}^T \otimes \boldsymbol{M}_{R^{(\ell)}}^T)\right)^{-1}, \tag{12}$$

where $\boldsymbol{I}^{(\ell)}$ is identity matrix with dimensions corresponding to the number of parameters in the ℓth layer and $\boldsymbol{M}_{Q^{(\ell)}} \otimes \boldsymbol{M}_{R^{(\ell)}}$ constitutes the eigenvectors of $\boldsymbol{Q}^{(\ell)} \otimes \boldsymbol{R}^{(\ell)}$, making it unitary and facilitating the representation of this identity. The LA and approximated posterior covariance via generalized Gauss–Newton and Kronecker-factored approximations, along with the eigendecomposition of the individual Kronecker factors, results in an efficient and scalable algorithm for Bayesian inference of neural networks.

3. Hierarchical Bayesian inferences and model plausibility

3.1. Bayesian neural networks model

We introduce the *BayesNN model* denoted by the abstract form $M(\boldsymbol{\phi})$. This representation encapsulates not only the network architecture but also the specifications of the prior of the parameters and the noise model utilized for constructing the likelihood. We thus categorize the BayesNN model parameters $\boldsymbol{\phi}$ into four distinct subsets, as follows:

(1) Network parameters $\boldsymbol{\theta} \in \mathbb{R}^P$. These encompass the set of weights and biases constituting a BayesNN, represented collectively as,

$$\boldsymbol{\theta} = [\boldsymbol{w}^{(1)}, \dots, \boldsymbol{w}^{(D)}, \boldsymbol{b}^{(1)}, \dots, \boldsymbol{b}^{(D)}]. \tag{13}$$

As discussed earlier, the primary objective of BayesNN training is to determine the probability distributions of the network parameters based on data.

(2) Inference hyper-parameters $\boldsymbol{\sigma} \in \mathbb{R}^H$. These parameters characterize the Bayesian inference setup and stem from the specifications of the prior and likelihood. More precisely, the prior on the weights and biases, denoted as $\pi_{pr}(\boldsymbol{\theta})$, can be parameterized by $\boldsymbol{\sigma}^{\text{pr}} \in \mathbb{R}^P$, such as mean and variance of a Gaussian prior. Additionally, the likelihood function incorporates noise hyper-parameters $\boldsymbol{\sigma}^{\text{noise}} \in \mathbb{R}^{N_D}$ arising from the chosen noise model. The inference hyper-parameters are a collection of prior and noise hyper-parameters,

$$\boldsymbol{\sigma} = [\boldsymbol{\sigma}^{\text{pr}}, \boldsymbol{\sigma}^{\text{noise}}]. \tag{14}$$

(3) Architecture hyper-parameters $\boldsymbol{\xi} \in \Xi$. These parameters represent the structural and organizational characteristics of the network, specifying the functional mapping between inputs and outputs. In the case of fully connected networks, the architecture hyper-parameters include the number of layers D, the number of neurons W and the activation function f in each layer (i.e., the type of non-linearity applied between layers),

$$\boldsymbol{\xi} = [W^{(1)}, \dots, W^{(D)}, f^{(1)}, \dots, f^{(D)}]. \tag{15}$$

Here, $D, W \in \mathbb{N}$ and $f \in \mathcal{F}$, where $\mathcal{F}$ denotes the function space adhering to properties expected of neural network activation functions, such as smoothness, continuity, and convexity. For more advanced models, like convolutional neural networks, the space of architecture hyper-parameters expands significantly, including the number and kernel size of convolutions, channels per convolutional layer, and the number of fully-connected layers, e.g., [31,62].

(4) Solution hyper-parameters $\chi \in \mathbb{R}^X$. Another class of parameters pertains to the selection of Bayesian inference solutions, encompassing choices like sampling algorithms and variational inference, along with the particulars of these algorithms, such as learning rate and weight initialization. In this work, we do not investigate the influence of solution hyperparameters on the predictive capability of the models. Hence, throughout the rest of this manuscript, the BayesNN model explicitly specifies the network's functional form, as well as the likelihood and prior specifications in Bayesian inference.

3.2. Model evidence

A common method for neural network model selection is based on their prediction performance on unseen testing datasets. However, this approach has been criticized for irreproducible outcomes [63], lack of robustness to small and uncertain data [49], and resulting in suboptimal architectures [64]. Instead, we advocate a principled model selection method based on Bayesian model plausibility, leveraging the entire dataset for model construction. Let $M(\boldsymbol{\phi})$ be a BayesNN model with its own set of model parameters $\boldsymbol{\phi} = [\boldsymbol{\theta}, \boldsymbol{\sigma}, \boldsymbol{\xi}]$. We rewrite the Bayesian inference (3), acknowledging known information on the inference and architecture hyper-parameters,

$$\pi_{post}(\boldsymbol{\theta}|\boldsymbol{D}, \boldsymbol{\sigma}, \boldsymbol{\xi}) = \frac{\pi_{like}(\boldsymbol{D}|\boldsymbol{\theta}, \boldsymbol{\sigma}, \boldsymbol{\xi})\pi_{pr}(\boldsymbol{\theta}|\boldsymbol{\sigma}, \boldsymbol{\xi})}{\pi_{evid}(\boldsymbol{D}|\boldsymbol{\sigma}, \boldsymbol{\xi})}. \tag{16}$$

The denominator of Bayes theorem (16) is the probability of observing data using model M, commonly referred to as the evidence PDF,

$$\pi_{evid}(\boldsymbol{D}|\boldsymbol{\sigma}, \boldsymbol{\xi}) = \int \pi_{like}(\boldsymbol{D}|\boldsymbol{\theta}, \boldsymbol{\sigma}, \boldsymbol{\xi})\, \pi_{pr}(\boldsymbol{\theta}|\boldsymbol{\sigma}, \boldsymbol{\xi})\, d\boldsymbol{\theta}. \tag{17}$$

In this setting, the evidence extends beyond its conventional role as a mere normalization factor in classical Bayesian inference, emerging as a pivotal component in our model selection framework. Specifically, by rearranging (3), the logarithm of evidence can be formulated as the posterior mean of the log-likelihood and the relative entropy between the prior and posterior distributions [65]. In essence, the evidence reflects a trade-off between the model fitting the observed data and the extent to which network parameters $\boldsymbol{\theta}$ are learned from that data. This characteristic renders the evidence a valuable measure for model comparison [22], particularly in BayesNN models, facilitating comparisons between different architectures, noise models, and prior formulations.

3.3. Hierarchical Bayesian inferences

Building upon MacKay's evidence framework [26], the following hierarchical Bayesian inferences are postulated, treating the evidence PDF at each level as a new likelihood function for subsequent inference:

- *Level 1.* Infer the network parameters $\boldsymbol{\theta}$,

$$\pi_{post}(\boldsymbol{\theta}|\boldsymbol{D}, \boldsymbol{\sigma}, \boldsymbol{\xi}) = \frac{\pi_{like}(\boldsymbol{D}|\boldsymbol{\theta}, \boldsymbol{\sigma}, \boldsymbol{\xi})\, \pi_{pr}(\boldsymbol{\theta}|\boldsymbol{\sigma}, \boldsymbol{\xi})}{\pi_{evid}(\boldsymbol{D}|\boldsymbol{\sigma}, \boldsymbol{\xi})}. \tag{18}$$

- *Level 2.* Infer the inference hyper-parameters $\boldsymbol{\sigma}$,

$$\pi_{post}(\boldsymbol{\sigma}|\boldsymbol{D}, \boldsymbol{\xi}) = \frac{\pi_{evid}(\boldsymbol{D}|\boldsymbol{\sigma}, \boldsymbol{\xi})\, \pi_{pr}(\boldsymbol{\sigma}|\boldsymbol{\xi})}{\pi_{evid}(\boldsymbol{D}|\boldsymbol{\xi})}. \tag{19}$$

- *Level 3.* Infer the architecture hyper-parameters $\boldsymbol{\xi}$,

$$\pi_{post}(\boldsymbol{\xi}|\boldsymbol{D}) = \frac{\pi_{evid}(\boldsymbol{D}|\boldsymbol{\xi})\, \pi_{pr}(\boldsymbol{\xi})}{\pi_{evid}(\boldsymbol{D})}. \tag{20}$$

Although leveraging *Level 1* and *Level 2* inferences in BayesNN has gained recent attention [47], the utilization of *Level 3* for neural network architecture selection has been relatively understudied. In particular, (20) facilitates the update of the modeler's prior belief in a BayesNN model $\pi_{pr}(\boldsymbol{\xi})$ based on data, yielding the posterior model plausibility $\rho = \pi_{post}(\boldsymbol{\xi}|\boldsymbol{D})$.

With the posterior PDFs of model parameters and hyperparameters in hand, the next step is to make prediction using the updated model. For a fixed architecture, this entails marginalizing over uncertainty at the first two levels to derive the prediction distribution,

$$\pi(\boldsymbol{u}_{\boldsymbol{\theta}}|\boldsymbol{D},\boldsymbol{\xi}) = \int \pi(\boldsymbol{u}_{\boldsymbol{\theta}}|\boldsymbol{\theta},\boldsymbol{\sigma},\boldsymbol{\xi})\,\pi(\boldsymbol{\theta},\boldsymbol{\sigma}|\boldsymbol{D},\boldsymbol{\xi})\,d\boldsymbol{\sigma}\,d\boldsymbol{\theta}, \tag{21}$$

where $\pi(\boldsymbol{\theta},\boldsymbol{\sigma}|\boldsymbol{D},\boldsymbol{\xi}) = \pi_{post}(\boldsymbol{\theta}|\boldsymbol{D},\boldsymbol{\sigma},\boldsymbol{\xi})\pi_{post}(\boldsymbol{\sigma}|\boldsymbol{D},\boldsymbol{\xi})$ and the evaluation of $\pi(\boldsymbol{u}_{\boldsymbol{\theta}}|\boldsymbol{\theta},\boldsymbol{\sigma},\boldsymbol{\xi})$ requires a single forward model evaluation at the desired inputs. Similarly, the predictions of multiple architectures are obtained by computing a weighted sum of (21), where the model plausibility ρ serves as the weight, known as Bayesian model averaging, e.g., [49,66,67].

Remark 1. In the hierarchical Bayesian inference of (18) and (19), there may appear to be a contradiction with Bayesian principles, as the prior for network parameters seems to be determined after observing the data. In other words, it might appear that the most probable value of the prior hyperparameters $\boldsymbol{\sigma}^{\text{pr}}$ is chosen first, followed by the utilization of the corresponding prior to infer $\boldsymbol{\theta}$. To clarify this apparent sequence, we recognize that the model prediction can be the integration over the ensemble of unknown network and inference hyperparameters collectively defining the prior. The true posterior of the network parameters used for model prediction in (21) is expressed as:

$$\pi_{post}(\boldsymbol{\theta}|\boldsymbol{D},\boldsymbol{\xi}) = \int \pi(\boldsymbol{\theta},\boldsymbol{\sigma}|\boldsymbol{D},\boldsymbol{\xi})\,d\boldsymbol{\sigma} = \int \pi_{post}(\boldsymbol{\theta}|\boldsymbol{D},\boldsymbol{\sigma},\boldsymbol{\xi})\,\pi_{post}(\boldsymbol{\sigma}|\boldsymbol{D},\boldsymbol{\xi})\,d\boldsymbol{\sigma}. \tag{22}$$

This expression suggests that the posterior of $\boldsymbol{\theta}$ is obtained by integrating over the posteriors for all values of $\boldsymbol{\sigma}$, each weighted by the probability of $\boldsymbol{\sigma}$ given the data. This mirrors the concept of Bayesian model averaging, where predictions from multiple BayesNN with different inference hyperparameters are combined, considering the associated uncertainty of each model. Considering $\pi_{post}(\boldsymbol{\sigma}|\boldsymbol{D},\boldsymbol{\xi})$ exhibiting a dominant peak at $\boldsymbol{\sigma}_{MAP}$, then the true posterior $\pi_{post}(\boldsymbol{\theta}|\boldsymbol{D},\boldsymbol{\xi})$ will be primarily governed $\pi_{post}(\boldsymbol{\theta}|\boldsymbol{D},\boldsymbol{\sigma},\boldsymbol{\xi})|_{\boldsymbol{\sigma}=\boldsymbol{\sigma}_{MAP}}$. Consequently, we are justified in using the following approximation [45],

$$\pi_{post}(\boldsymbol{\theta}|\boldsymbol{D},\boldsymbol{\xi}) \approx \pi_{post}(\boldsymbol{\theta}|\boldsymbol{D},\boldsymbol{\sigma},\boldsymbol{\xi})|_{\boldsymbol{\sigma}=\boldsymbol{\sigma}_{MAP}}. \tag{23}$$

Therefore, employing only the most probable prior hyperparameter is a valid approximation for posterior evaluation. We note that the Bayesian model averaging can be extended to combine predictions from multiple network architectures, with each model weighted according to its posterior probability based on the data according to (20).

Numerical solution. The solution to hierarchical Bayesian inferences involves a multi-step procedure. Specifically, *Level 1* and *Level 2* inferences comprise an iterative scheme for the online determination of network parameters and inference hyperparameters. In each step, the network parameters $\boldsymbol{\theta}$ are inferred with fixed values of the inference hyperparameters $\boldsymbol{\sigma}$ using (18). Subsequently, these parameters are utilized to update $\boldsymbol{\sigma}$ as per (19), maximizing the corresponding evidence $\pi_{evid}(\boldsymbol{D}|\boldsymbol{\sigma},\boldsymbol{\xi})$. Upon completion of the network training and hyperparameters inference, the evidence $\pi_{evid}(\boldsymbol{D}|\boldsymbol{\xi})$ is employed to compute posterior model plausibility in *Level 3*. Using LA to the posterior and the prior and noise model considered in Section 2, the process of network training and hyperparameter determination entails computing their corresponding MAP estimates and the posterior covariance. In the absence of prior information on the inference hyperparameters and the non-Gaussian form of likelihood, we adopt uniform priors for $\ln(\sigma^{\text{pr}^2})$ and $\ln(\sigma^{\text{noise}^2})$, in accordance with the recommendation of [26,45]. This prior facilitates exploration across a broad spectrum of hyperparameters and ensures cancellation when comparing different models. Taking $\boldsymbol{\sigma} = [\ln(\sigma^{\text{noise}^2}), \ln(\sigma^{\text{pr}^2})]$, the evidence for updating inference hyperparameters is expressed as,

$$\begin{aligned}\ln \pi_{evid}(\boldsymbol{D}|\boldsymbol{\sigma},\boldsymbol{\xi}) &\approx \ln \pi_{like}(\boldsymbol{D}|\boldsymbol{\theta},\boldsymbol{\sigma},\boldsymbol{\xi})|_{\boldsymbol{\theta}=\boldsymbol{\theta}_{MAP}} + \ln \pi_{pr}(\boldsymbol{\theta}|\boldsymbol{\sigma},\boldsymbol{\xi})|_{\boldsymbol{\theta}=\boldsymbol{\theta}_{MAP}} - \ln \pi_{post}(\boldsymbol{\theta}|\boldsymbol{D},\boldsymbol{\sigma},\boldsymbol{\xi})|_{\boldsymbol{\theta}=\boldsymbol{\theta}_{MAP}} \\ &= -\frac{1}{2\ln(\sigma^{\text{noise}^2})}\sum_{j=1}^{N_D}(\boldsymbol{u}_D^j - \boldsymbol{u}_{\boldsymbol{\theta}_{MAP}}(\boldsymbol{x}_D^j))^2 - \frac{N_D}{2}\ln(2\pi\ln(\sigma^{\text{noise}^2})) - \frac{1}{2\ln(\sigma^{\text{pr}^2})}\|\boldsymbol{\theta}_{MAP}\|^2 - \frac{P}{2}\ln(2\pi\ln(\sigma^{\text{pr}^2})) - \frac{1}{2}\ln\det\left(\frac{1}{2\pi}\boldsymbol{H}\right),\end{aligned} \tag{24}$$

where $\|\cdot\|$ denotes the Euclidean norm. Utilizing a Gaussian approximation of $\pi_{post}(\boldsymbol{\sigma}|\boldsymbol{D},\boldsymbol{\xi})$, the MAP estimate of the inference hyperparameters in (19) is obtained by maximizing the evidence in (24). The posterior variances of $\ln(\sigma^{\text{noise}^2})$ and $\ln(\sigma^{\text{pr}^2})$ are derived by differentiating (24) twice, following the derivations provided in [45], as

$$\sigma^2_{\ln(\sigma^{\text{pr}^2})} \approx \frac{2\left(\ln(\sigma^{\text{pr}^2}_{MAP})\right)^2}{P - \frac{1}{\ln(\sigma^{\text{pr}^2}_{MAP})}\text{tr}(\boldsymbol{H}^{-1})},\ \sigma^2_{\ln(\sigma^{\text{noise}^2})} \approx \frac{2\left(\ln(\sigma^{\text{pr}^2}_{MAP})\right)^2}{N_D - P + \frac{1}{\ln(\sigma^{\text{pr}^2}_{MAP})}\text{tr}(\boldsymbol{H}^{-1})}. \tag{25}$$

Using (25), the model evidence for evaluating the posterior model plausibility in *Level 3* is expressed as,

$$\pi_{evid}(\boldsymbol{D}|\boldsymbol{\xi}) = \int \pi_{evid}(\boldsymbol{D}|\boldsymbol{\sigma},\boldsymbol{\xi})\,\pi_{pr}(\boldsymbol{\sigma}|\boldsymbol{\xi})\,d\boldsymbol{\sigma} \approx \pi_{evid}(\boldsymbol{D}|\boldsymbol{\sigma},\boldsymbol{\xi})|_{\boldsymbol{\sigma}=\boldsymbol{\sigma}_{MAP}}\,\pi_{pr}(\boldsymbol{\sigma}|\boldsymbol{\xi})|_{\boldsymbol{\sigma}=\boldsymbol{\sigma}_{MAP}}\left(2\pi\sigma_{\ln(\sigma^{\text{pr}^2})}\sigma_{\ln(\sigma^{\text{noise}^2})}\right), \tag{26}$$

where $\pi_{evid}(\boldsymbol{D}|\boldsymbol{\sigma},\boldsymbol{\xi})|_{\boldsymbol{\sigma}=\boldsymbol{\sigma}_{MAP}}$ is evaluated using (24).

The iterative updating network parameters, inference hyperparameters, and computing model plausibility is computationally expensive due to calculations involving the determinant of the Hessian in (24) and the trace of its inverse in (26) for high-dimensional

parameters. However, we leverage the available eigenvalues of the Kronecker-factored matrices described in Section 2.2 for efficient computation of the determinant and trace,

$$\det(\boldsymbol{H}) = \prod_{\ell}\prod_{ij} q_i^{(\ell)} r_j^{(\ell)} + (\sigma^{\mathrm{pr}(\ell)})^2, \quad \mathrm{tr}(\boldsymbol{H}) = \sum_{\ell}\sum_{ij} q_i^{(\ell)} r_j^{(\ell)} + (\sigma^{\mathrm{pr}(\ell)})^2. \tag{27}$$

4. A framework for surrogate model discovery

The main focus of this study is on surrogate modeling, specifically utilizing BayesNN models as low-computational-cost approximations for solutions of high-fidelity physics-based simulations. Beyond addressing uncertainties arising from limited and sparse data in high-fidelity simulations, credible surrogate models must capture the overall hierarchy of interactions across multiple scales inherent in physics-based simulations [68]. We argue that Bayesian training of neural networks is well-suited for this purpose, as it allows for more complex models with an adequate number of weights to capture the multiscale structure of physics-based simulation. This stands in contrast to maximum likelihood training, where an excessive number of parameters may lead to overfitting and compromise the models' generalization. Despite these merits, challenges persist in surrogate modeling using BayesNN, particularly in identifying the right model within an enormous space of potential architectures and rigorous methodologies for assessing the accuracy and reliability of surrogates under uncertainty.

In formulating the surrogate modeling problem, let $\mathcal{M}$ be a set of different BayesNN models,

$$\mathcal{M} = \left\{ M_1(\boldsymbol{\phi}_1), M_2(\boldsymbol{\phi}_2), \dots, M_K(\boldsymbol{\phi}_K) \right\}, \tag{28}$$

where each model $M_i,\ i = 1, 2, \dots, K$ has its own set of model parameters $\boldsymbol{\phi}_i = [\theta_i, \boldsymbol{\sigma}_i, \boldsymbol{\xi}_i]$. The high-fidelity data $\boldsymbol{D}$ is obtained from the scenario $\mathcal{S}$ that encapsulates features of the high-fidelity simulation such as domain, boundary conditions, and spatial location and frequency from which data is gathered. Later we will argue the importance of considering a hierarchy of scenarios to rigorously assess the validity and predictive reliability of surrogate models. The training process for each model involves hierarchical Bayesian inferences (18)–(20), where we recognize known information about the set $\mathcal{M}$ and scenario $\mathcal{S}$ in the probability distributions, i.e., $\pi_{post}(\boldsymbol{\theta}|\boldsymbol{D}, \boldsymbol{\sigma}, \boldsymbol{\xi}, \mathcal{M}, \mathcal{S})$. In this context, the denominator of (20) serves to normalize the discrete probability components over the models in the set $\mathcal{M}$,

$$\pi(\boldsymbol{D}|\mathcal{M}, \mathcal{S}) = \sum_{i=1}^{K} \pi_{evid}(\boldsymbol{D}|\boldsymbol{\xi}_i, \mathcal{M}, \mathcal{S}) \pi_{pr}(\boldsymbol{\xi}_i|\mathcal{M}), \tag{29}$$

resulting in $\sum_{i=1}^{K} \rho_i = 1$, with $\rho_i = \pi_{post}(\boldsymbol{\xi}_i|\boldsymbol{D})$. Hence, within the set of competing models in $\mathcal{M}$ and for a given dataset $\boldsymbol{D}$, the model boasting the highest ρ_i is deemed the most plausible model [22,65]. However, the model evidence alone does not ensure the generalizability and prediction reliability of the model, indicating the need for additional steps for credibility assessment and model validation.

4.1. Occam-plausibility algorithm for surrogates

Building on our previous framework for validation and selection of physics-based models [22,23,43], we proposed Occam-Plausibility Algorithm for Surrogate models (OPAL-surrogate), a systematic strategy for discovering the simplest ("best") credible surrogate model shown in Fig. 1. The key enabler to overcome the enormous space of potential BayesNN models is adaptive modeling by replacing discrete choices with continuous parameterization. The OPAL-surrogate involves the following steps:

1. **Initialization.** Construct an initial set of possible surrogate models $\mathcal{M}$ in (28) and acquired training data $\boldsymbol{D}$ from a scenario $\mathcal{S}^{(s)}$ with setting $s = 1$.
2. **Occam Categories.** Partition the models in $\mathcal{M}$ based on model complexity measures into *Occam categories*. The simplest models are placed in Category 1, and the most complex models are designated in the last category. We, therefore, produce a collection of subsets,

$$\hat{\mathcal{M}}^{(l)} = \left\{ M_1^l(\boldsymbol{\phi}_1^l), M_2^l(\boldsymbol{\phi}_2^l), \dots, M_{K_l}^l(\boldsymbol{\phi}_{K_l}^l) \right\}, \quad l = 1, 2 \cdots, H, \tag{30}$$

   where l is the category and $K_l < K$. A commonly used complexity measure is the number of model parameters $\boldsymbol{\phi}$, although other measures can also be considered, especially when implementing OPAL-surrogate to multiple classes of surrogate models. The categorization of models depends on factors such as the size of the initial model set and the available computational resources.
3. **Occam Step and Model Plausibility.** Start with the first Occam Category, $l = 1$, employ $\boldsymbol{D}$ to identify the plausible model in Category l, denoted by $M_I^l(\boldsymbol{\phi}_I^l)$. This process is formalized as,

$$\boldsymbol{\xi}_I^l = \underset{\boldsymbol{\xi}_k^l \in \bar{\Xi}^l}{\mathrm{argmax}} \ \{\rho_k^l\}_{k=1}^{R_l}, \tag{31}$$

   where $\rho_k^l = \pi_{post}(\boldsymbol{\xi}_k^l|\boldsymbol{D}, \hat{\mathcal{M}}^{(l)}, \{\mathcal{S}^{(s)}\}_{s=1}^{S})$ represents the Bayesian posterior plausibility for the kth model in the set $\hat{\mathcal{M}}^{(l)}$, given the high-fidelity data $\boldsymbol{D}$ obtained from a collection of scenarios $\mathcal{S}^{(1)}, \dots, \mathcal{S}^{(S)}$.

Except for rare instances like model selection among a limited space of architectures $\bar{\Xi}^l$, such as in [58], discrete optimization in (31) that requires Bayesian inference across all possible models to evaluate ρ_i^l, becomes computationally prohibitive. Instead, one can adopt a uniform prior on all (or a subset of) architecture hyper-parameters $\pi_{pr}(\boldsymbol{\xi}_i|\hat{\mathcal{M}}^{(l)})$, and employ adaptive modeling by replacing discrete model choices with continuous parameterization. This transformation turns the optimization problem in (31) into a continuous maximization of the evidence $\pi_{evid}(\boldsymbol{D}|\boldsymbol{\xi}_i, \hat{\mathcal{M}}^{(l)}, \{S^{(s)}\}_{s=1}^S)$, facilitating efficient model discovery. Solving the possibly non-convex optimization problem using pure machine learning techniques involves leveraging advanced neural search algorithms [28–30,34]. However, inspired by [68,69], our adaptive surrogate modeling takes a distinctive perspective to ensure the BayesNN approximator effectively captures the underlying multiscale structure of the high-fidelity physics-based simulations. To realize this objective, we propose a sequential addition of fully connected layers with sufficiently large widths, guided by the model evidence value, followed by the elimination of irrelevant weights. This strategic approach necessitates an effective network sparsification method to reveal sparsity patterns associated with the inherent multiscale structure encoded in high-fidelity data. One of such methods is described in Section 4.2.

4. **Credibility Assessment.** Examine the validity of the most plausible model $M_I^l(\boldsymbol{\phi}_I^l)$, by subjecting it to *leave-out cross-validation test*. Accordingly, divide the training data $\boldsymbol{D}$ into N_v leave-out subsets,

$$\{\boldsymbol{D}_{LO}^{(n)}\}_{n=1}^{N_v} \quad \text{with} \quad \boldsymbol{D}_{LO}^{(n)} = \boldsymbol{D} \setminus \boldsymbol{D}^{(n)}. \tag{32}$$

For each subset of data, train $M_I^l(\boldsymbol{\phi}_I^l)$ with $\boldsymbol{D}^{(n)}$ using the Bayesian inference (18) and (19). Use the posteriors of the network parameters and inference hyper-parameters to evaluate the model output $\boldsymbol{u}_{\boldsymbol{\theta}}(\boldsymbol{x}_{D_{LO}^{(n)}})$ associated with $\boldsymbol{D}_{LO}^{(n)}$. Compare a proper distance measure $\mathbb{d}(\cdot,\cdot)$ between the model output and the leave-out data set with a given accuracy tolerance TOL. For $n = 1, \ldots, N_v$,

$$\begin{aligned}
&\pi_{post}(\boldsymbol{\theta}_I^l|\boldsymbol{D}^{(n)}, \boldsymbol{\sigma}_I^l, \boldsymbol{\xi}_I^l, \hat{\mathcal{M}}^{(l)}, \{S^{(s)}\}_{s=1}^S) = \frac{\pi_{like}(\boldsymbol{D}^{(n)}|\boldsymbol{\theta}_I^l, \boldsymbol{\sigma}_I^l, \boldsymbol{\xi}_I^l, \hat{\mathcal{M}}^{(l)}, \{S^{(s)}\}_{s=1}^S)}{\pi_{evid}(\boldsymbol{D}^{(n)}|\boldsymbol{\sigma}_I^l, \boldsymbol{\xi}_I^l, \hat{\mathcal{M}}^{(l)}, \{S^{(s)}\}_{s=1}^S)} \pi_{pr}(\boldsymbol{\theta}_I^l|\boldsymbol{\sigma}_I^l, \boldsymbol{\xi}_I^l),\\
&\pi_{post}(\boldsymbol{\sigma}_I^l|\boldsymbol{D}^{(n)}, \boldsymbol{\xi}_I^l, \hat{\mathcal{M}}^{(l)}, \{S^{(s)}\}_{s=1}^S) = \frac{\pi_{evid}(\boldsymbol{D}^{(n)}|\boldsymbol{\sigma}_I^l, \boldsymbol{\xi}_I^l, \hat{\mathcal{M}}^{(l)}, \{S^{(s)}\}_{s=1}^S)}{\pi_{evid}(\boldsymbol{D}^{(n)}|\boldsymbol{\xi}_I^l, \hat{\mathcal{M}}^{(l)}, \{S^{(s)}\}_{s=1}^S)} \pi_{pr}(\boldsymbol{\sigma}_I^l|\boldsymbol{\xi}_I^l),\\
&\mathbb{d}^{(n)}\left(\pi(\boldsymbol{u}_{\boldsymbol{D}}(\boldsymbol{x}_{D_{LO}^{(n)}})),\ \pi(\boldsymbol{u}_{\boldsymbol{\theta}}(\boldsymbol{x}_{D_{LO}^{(n)}}))\right) < TOL.
\end{aligned} \tag{33}$$

If the inequality $(33)_3$ holds for all N_v, the model M_I^l is considered as "not invalid" and identified as the *best credible surrogate model* for the given training data. This designation is made under the recognition that additional data and information could potentially falsify a model initially presumed to be valid. Once the surrogate model is determined, the parameters are updated using the entire training set. This involves substituting $\boldsymbol{D}^{(n)}$ with $\boldsymbol{D}$ in $(33)_{1,2}$, and utilizing the inferred network parameters $\pi_{post}(\boldsymbol{\theta}_I^l|\boldsymbol{D}, \boldsymbol{\sigma}_I^l, \boldsymbol{\xi}_I^l, \hat{\mathcal{M}}^{(l)}, \{S^{(s)}\}_{s=1}^S)$ for making predictions.

5. **Scenario Design via Active Learning.** If additional computational resources allow for further data generation, the model $M_I^l(\boldsymbol{\phi}_I^l)$ is utilized to design the next scenario $s \leftarrow s+1$ to augment $\boldsymbol{D}$ with more effective data, following the active learning approach. In cases where surrogate models are utilized for interpolation, this involves leveraging optimal experimental design methods, such as [70–72], which take a decision-theoretic approach to optimize features of the high-fidelity simulation scenario by maximizing an information gain metric as expected utility. However, complications arise in extrapolation when the surrogate model aims to predict *unobservable QoI* beyond the scope of high-fidelity simulation. This poses the formidable challenge of designing a scenario to obtain observational data that reflects the structure of the QoI [23,73] and further discussed in Remark 2 below.
6. **Iteration and Refinements.** If the model M_I^l is invalid, OPAL returns to the next Occam category $l \leftarrow l+1$ and repeats Steps 3 and 4 until identifying a "not invalid" model and possibly augmenting training data in Step 5. In case all BayesNN models in $\mathcal{M}$ are found to be invalid, it is necessary to enlarge the initial model set.

Remark 2. While the OPAL-surrogate provides an adaptive strategy for discovering neural network-based surrogate models, it emphasizes the vital role of integrating domain expert knowledge into the modeling process. The efficacy of this framework relies on various subjective decisions that the modeler must tailor to a specific problem. These decisions involve defining the initial model set, selecting appropriate complexity measures and categorization methods, establishing the prior distribution for architecture hyperparameters, defining metrics for credibility assessment, and choosing the utility function in scenario design. Of particular significance are the credibility assessment (step 4) and the choice of validation tolerance. In surrogate modeling scenarios, such as those illustrated in the numerical examples in Section 5, where the objective is to identify architectures for a specific class of neural network models, a convergence study perspective may be considered. Instead of using a fixed tolerance, the model is selected within the Occam Category, where predictive performance diminishes with increasing complexity in the higher category. However, when extending the application of the OPAL-surrogate, for instance, to identify predictive models among diverse classes of neural operators and varied architectures, the convergence perspective becomes highly sensitive to categorization. Therefore, it is more appropriate to consider a user-defined, problem-specific tolerance on the validation observables, indicating the acceptable level of prediction errors in the surrogate model. This approach also allows for the possibility of the OPAL-surrogate rejecting all models in the initial set, prompting the identification of new potential models to achieve acceptable prediction accuracy, rather than solely relying on models within the initial set.

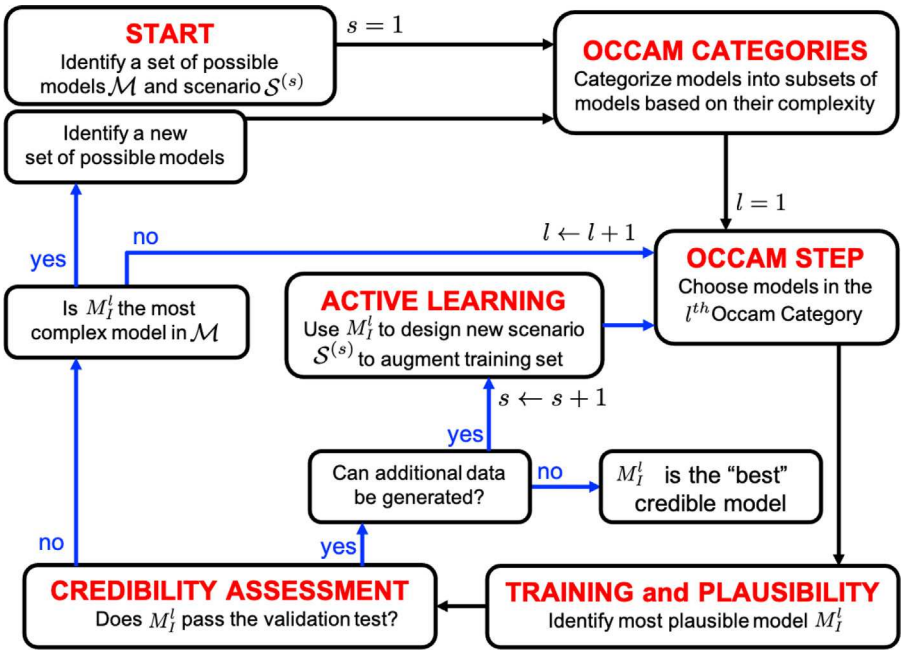

Fig. 1. The Occam Plausibility Algorithm for surrogate models (OPAL-surrogate): Commencing with an initial set of potential models, the models are categorized based on measures of complexity. Starting with the first Occam Category, the most plausible model is determined and undergoes the validation test for credibility assessment. If additional data can be acquired, the validated model guides the design of new scenarios to expand the training set and iteratively re-select new models in response to the updated dataset. The model that passes the validation test is considered the "best" credible model.

Remark 3. A formidable challenge in scientific prediction is the extrapolation beyond the range of available data to predict unobservable QoIs. Section 5 presents a numerical example illustrating this situation, where the surrogate's objective is to make predictions on a larger domain inaccessible to high-fidelity simulation. The formalization of surrogate modeling, grounded in the concept of a "prediction pyramid" [20,22], is pivotal in augmenting the reliability of extrapolation predictions. At the base of the prediction pyramid is the pre-training scenario $\mathcal{S}_c$, with a lower computational cost of the high-fidelity simulation. It facilitates the generation of a large set of pre-training data $\boldsymbol{D}_c$, contributing to establishing a meaningful prior distribution on the network parameters. Ascending the pyramid is the training scenario $\mathcal{S}$, involving more complex high-fidelity simulations that provide training dataset $\boldsymbol{D}$, used to inform and test the surrogate model's trustworthiness and prediction reliability. At the top of the pyramid is the prediction scenario $\mathcal{S}_p$, the most complex scenario where conducting high-fidelity simulation is practically impossible. The surrogate model prediction relies on the extrapolation of training data and the design of a meaningful training scenario that accurately captures the features of the QoIs in the prediction scenario [23,73].

4.2. Network sparsification strategy

As outlined in Section 4.1, an effective sparsification is needed to ensure the surrogate model captures the inherent multiscale structure within high-fidelity physical simulations. The conventional method involves pruning by setting tentative network parameters to zero, followed by training the new model to accept the pruning based on a performance measure. However, explicitly pruning one parameter at a time becomes computationally prohibitive for large networks [26]. In contrast, we seek to automatically assess the relevance of network parameters, preserving those deemed relevant to enhance the surrogate model's predictive performance. Here, we advocate exploiting sparsity-enforcing priors to eliminate irrelevant parameters of BayesNN, e.g., [74,75]. From a Bayesian perspective, such a prior reflects our belief that certain parameters are less likely to be relevant than others. This leads us to anticipate that these parameters will be centered around a specific value, with penalization for deviations from this mode. Employing the maximum entropy principle for constructing prior probability distributions [76], we obtain a Laplace distribution by imposing constraints on both the mean of the network parameters and their absolute deviation (L1 norm) from the mean, expressed as,

$$\pi_{pr}(\boldsymbol{\theta}) = \prod_{i=1}^{P} \frac{1}{\beta} \exp\left[-\frac{|\theta_i - \bar{\theta}_i|}{\beta}\right], \tag{34}$$

where $\beta > 0$ is the scale parameter. The network sparsification method involves applying a Laplace prior to the network parameters, followed by magnitude-based thresholding. Parameters with sufficiently small magnitudes, denoted as $|\theta_i| < TOL_\theta$, are considered irrelevant and subsequently removed from the model. The level of aggressiveness in parameter removal is controlled to maximize the model evidence, achieved by incrementally increasing the threshold TOL_θ until the evidence value of the resulting sparsed network begins to decline.

Discovering the category-wise plausible model. We propose the following strategy for implementing Step 3 of OPAL-surrogate across a wide range of possible BayesianNN models. In this context, we characterize fully connected neural network architectures by their depth D (number of layers), width W (number of neurons in each layer), and each layer's activation function, employing a uniform prior on all architecture hyper-parameters:

1. Within each Occam category, identify a fully connected network architecture with the smallest depth and largest width. Among different choices of activation functions for the layer, select the one that yields the highest evidence $\pi_{evid}(\boldsymbol{D}|\boldsymbol{\xi}^l_{(\cdot)}, \hat{\mathcal{M}}^l, \mathcal{S})$.
2. Sequentially add a fully connected layer and select the corresponding activation function based on evidence value.
3. Among all fully connected networks with different layers $M^l_{F(D,W)}$, choose the one with the largest model evidence and subject that model to appropriate network sparsification. The resulting network, with eliminated irrelevant neural connections (corresponding weight and bias parameters), is deemed the plausible network M^l_I in this category.

The proposed strategy combines bottom-up (adding layers) and top-down (removing connections) approaches, ensuring the retention of essential parameters and unveiling the sparsity pattern needed to capture multiscale interactions within a high-fidelity dataset.

5. Numerical results

This section outlines numerical experiments on the OPAL-surrogate implementation for two high-fidelity physical simulations for problems in solid mechanics and computational fluid dynamics. The first application focuses on the elastic deformation of porous materials through which we explore hierarchical Bayesian inference and the proposed network sparsification method. OPAL-surrogate is then applied to identify the BayesNN surrogate model, ensuring accuracy and reliability in predicting strain energy as QoI and facilitating forward uncertainty quantification for material systems with domain sizes beyond the capacity of high-fidelity physical simulations. The second application focuses on the direct numerical simulation of turbulent combustion flow. Specifically, we apply OPAL-surrogate to determine BayesNN models for the interpolation prediction of combustion dynamics in shear-induced ablation of solid fuels within hybrid rocket motors. The implementation of Bayesian neural networks is built upon the Laplace Redux library [77], finite element solutions are conducted using the FEniCS library [78], and direct numerical simulations of turbulent combustion flow are performed using the ABLATE library [79].

5.1. Elasticity in porous materials with random microstructure

Our first application focuses on the deformation of porous silica aerogel, a high-performance insulation material for net-zero buildings, e.g., [37,80,81]. The high-fidelity model is characterized by a stochastic partial differential equation (PDE) representing the elastic deformation of the two-phase material. The primary model parameter governing the deformation behavior is the Young's modulus of the solid aerogel phase E_s. This model entails a stochastic microstructure indicator field, with samples derived from microstructural images of silica aerogel obtained from a lattice Boltzmann simulation of the foaming process; for further details, refer to [82]. The finite element solution of the PDE utilizes a uniformly fine mesh to resolve the microstructure patterns accurately. The model output is defined as the stochastic strain energy of the material system,

$$u_{\boldsymbol{D}} = \int_{\Omega} \boldsymbol{T} : \boldsymbol{E}\, d\boldsymbol{y}, \tag{35}$$

where $\boldsymbol{T}$ is the Cauchy stress and $\boldsymbol{E}$ is the strain tensors.

The prediction scenario $\mathcal{S}_p$ is illustrated at the top of Fig. 2 and involves an aerogel component with a characteristic length of $L = 297.5$ μm and applied traction $\boldsymbol{t} = (70.71, 70.71)$ N/m at the rightmost boundary. Given the computational demands associated with resolving microstructure patterns using a fine mesh, performing the high-fidelity simulation in this scenario is deemed computationally prohibitive. The objective of surrogate modeling is to approximate the solution of the high-fidelity simulation in $\mathcal{S}_p$ and enable the quantification of uncertainty in the strain energy, i.e., *unobservable QoI*, stemming from the stochastic microstructure and uncertainties in elasticity parameters. We thus adhere to the notation of the prediction pyramid, as elucidated in Remark 3 in Section 4.1, and adopt a hierarchy of scenarios for conducting high-fidelity simulations to generate data. As depicted in Fig. 2, the pre-training scenario $\mathcal{S}_c$ entails small domains subjected to uniaxial loads, while the training scenario $\mathcal{S}$ encompasses larger domains specifically designed to capture physical features of the QoI in the prediction scenario, such as stress localization and loading conditions. The pre-training dataset $\boldsymbol{D}_c$ comprises strain energies computed from 22 equidistant domain sizes within the range of $L = [17.5, 113.75]$ μm, 10 equidistant elastic moduli within $E_s = [100, 115]$ MPa, and 10 realizations of the aerogel microstructure patterns. The training set $\boldsymbol{D}$ includes strain energies from high-fidelity simulations obtained at 15 domain sizes within $L = [122.5, 183.75]$ μm, 10 elastic moduli, and 10 microstructure realizations, resulting in a total of 3800 data points in the pre-training and training sets. To assess extrapolation predictions, we consider the leave-out set $\boldsymbol{D}_{LO}$ containing data points with domain sizes $L_{LO} = \{179.4, 183.8\}$ μm. As the validation observables, we take the product of the strain energy and Young's modulus at a specific domain length L_{LO}, evaluated using the high-fidelity simulations and surrogate models, expressed as,

$$\mathcal{Z}_D = \int_{E_s^{low}}^{E_s^{up}} u_{\boldsymbol{D}}|_{L=L_{LO}}\, dE, \qquad \mathcal{Z}_M = \int_{E_s^{low}}^{E_s^{up}} u_{\theta}|_{L=L_{LO}}\, dE, \tag{36}$$

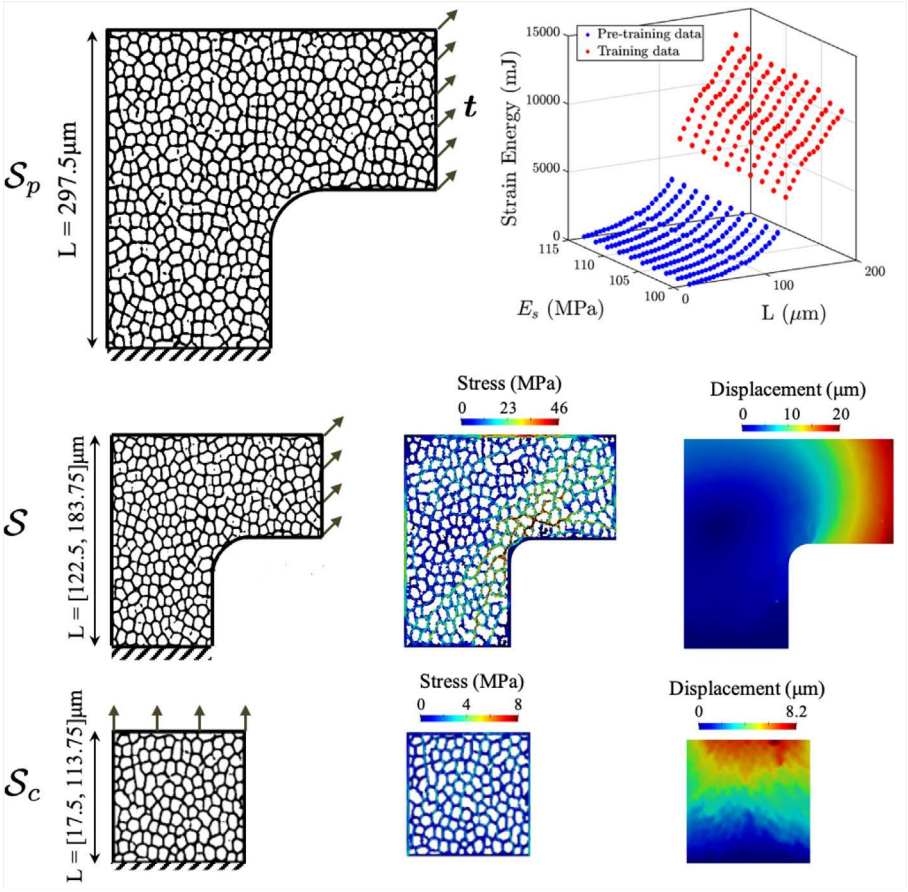

Fig. 2. Illustrations of the hierarchy of scenarios and high-fidelity data considered for the surrogate modeling of the elasticity problem: The pre-training scenario S_c consists of smaller domain sizes and under uniaxial loading conditions, while the training scenario S captures the mechanical features of the target prediction over domain sizes affordable for the high-fidelity model. The prediction scenario S_p involves an aerogel insulation component with a size of $L = 297.5$ μm, and unobservable QoI is the strain energy.

where $E_s^{low} = 100$ MPa and $E_s^{up} = 115$ MPa are the range of values for the elastic modulus within the training set. Given the stochastic nature of both observables, we employ two validation measures [83], the normalized Kullback–Leibler divergence by the Shannon entropy,

$$\mathrm{d}_{DKL} = \frac{\mathcal{D}_{\mathrm{KL}}\left(\pi(\mathcal{Z}_D), \pi(\mathcal{Z}_M)\right)}{\mathcal{H}\left(\pi(\mathcal{Z}_D)\right)}, \tag{37}$$

and discrepancies between cumulative distribution functions,

$$\mathrm{d}_{CDF} = \int_{-\infty}^{+\infty} |\Phi(\mathcal{Z}_D) - \Phi(\mathcal{Z}_M)| d\mathcal{Z}. \tag{38}$$

5.1.1. Illustrative 1D example

Prior to applying the OPAL-surrogate framework, we conduct a preliminary exploration of the methodologies detailed in Section 3 using a simple 1D example. In this example, the training data $\boldsymbol{D}$ consists of the strain energy calculations with $E_s = 100$ MPa over the domain sizes of both pre-training and training sets outlined in the preceding section.

Model plausibility vs. network architecture. Fig. 3 illustrates the log-evidence $\ln \pi_{evid}(\boldsymbol{D}|\boldsymbol{\xi}_i)$ for various fully connected neural networks with different depths D (number of layers), widths W (number of neurons in each layer), and activation functions. Considering uniform prior probabilities $\pi_{pr}(\boldsymbol{\xi}_i)$ for each model, the log-evidence values are equivalent to the posterior model plausibilities ρ_i. As depicted in these plots, a specific range of architectural hyperparameters leads to higher plausibilities, implying that the corresponding models have a higher probability of accurately representing the dataset. However, overly simplistic models, characterized by smaller D and W, exhibit limited predictive power within the dataset space. Conversely, excessively complex models with larger network parameters can capture a broader range of possible observations with low-confidence predictions within the dataset space. These aspects are illustrated in Fig. 4, depicting the mean and 95% credible interval (CI) uncertainty in predictions

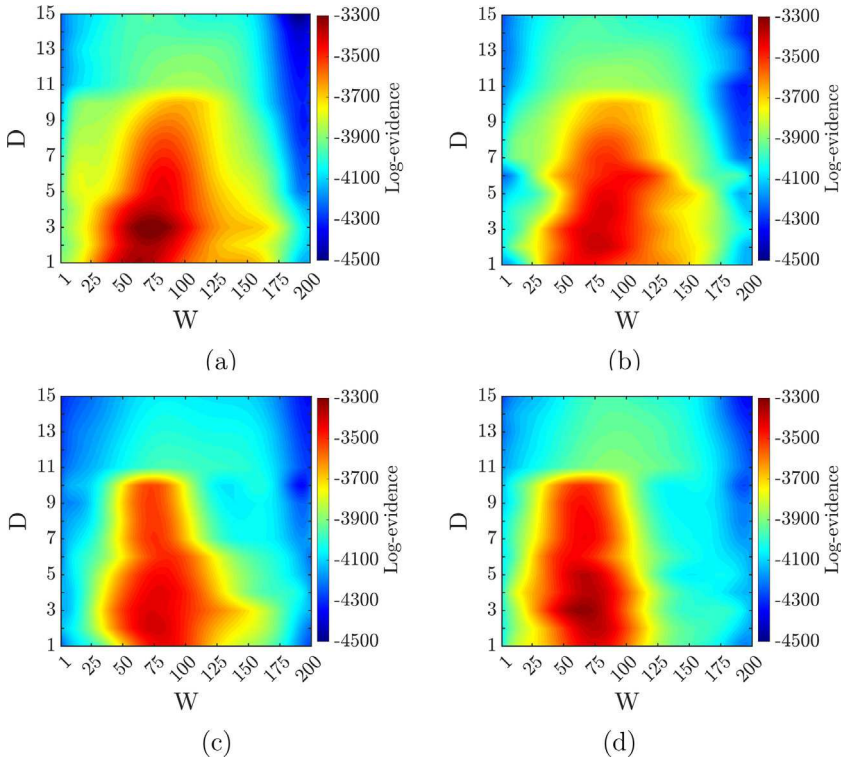

Fig. 3. Illustrative 1D example: The log-evidence $\ln \pi_{evid}(\boldsymbol{D}|\xi_i)$, corresponding to posterior model plausibilities ρ_i under the uniform prior $\pi_{pr}(\xi_i)$ assumption, for fully connected BayesNN models with varying number of layers D and neurons in each layer W, and activation functions: (a) *Tanh*, (b) *Leaky ReLU*, (c) *Sigmoid*, and (d) *ReLU*.

from BayesNN models with *Tanh* activation functions in comparison to the training data. As anticipated, owing to the extrapolation constraints inherent in data-driven surrogate models, model predictions consistently exhibit higher uncertainty beyond the data range compared to within the training data, across all architectures depicted in this figure. Models characterized by higher log evidence (Fig. 4(a) and (b)) present more accurate and reliable approximations for both interpolation and extrapolation relative to those with lower log evidence (Fig. 4(c) and (d)). For example, at $L = 150$ (μm) and $L = 250$ (μm), the standard deviations of model predictions for the model in Fig. 4(b) are 153.65 and 230.25, respectively, whereas, for the model depicted in Fig. 4(d), these values are 192.5 and 331.25. It is crucial to emphasize that model plausibility alone does not guarantee prediction credibility, underscoring the importance of validation tests, as discussed in Section 4.1, to rigorously assess the robustness of surrogate model predictions.

Network sparsification. Fig. 5 demonstrates the effectiveness of the sparsification method detailed in Section 4.2 in enhancing model plausibility by eliminating irrelevant network parameters. Gradually increasing TOL_θ initially increases the model evidence, followed by a decline due to excessive sparsification. The reported threshold values correspond to the maximum model evidence. In Fig. 5(a,b), the fully connected network exhibits a plateau in extrapolation with significant uncertainty. Sparsification reduces uncertainty (13.23% in prediction variance) by eliminating irrelevant parameters, albeit with limited enhancement in prediction accuracy. Conversely, Fig. 5(c,d) shows that network sparsification improves accuracy in extrapolation predictions by maintaining the trend in training data for larger values of L. While these figures depict two representative cases, additional experiments indicate that the proposed network sparsification generally enhances both accuracy and reliability in BayesNN models.

5.1.2. OPAL-surrogate demonstration

This section presents the results of identifying the credible surrogate model for predicting the unobservable QoI, the strain energy of the prediction scenario S_p, in the elasticity problem, as illustrated in Fig. 2. For training the surrogate models and computing the evidence and plausibility, we first infer network parameters using pre-training data $\boldsymbol{D}_c$ based on (18) and (19). The resulting posterior serves as the prior for network parameters in hierarchical Bayesian inferences using training data $\boldsymbol{D}$ to evaluate posterior model plausibility.

Determining the initial model set $\mathcal{M}$ and categorization. In defining a large model space, we use the model evidence (model plausibility) of BayesNNs with a single layer ($D = 1$), varying widths ($W = [1, 1200]$), and four activation functions, as shown in Fig. 6. We set the upper width limit at $W = 600$ in the initial model set $\mathcal{M}$, representing a 10% increase over the peak average log-evidences for different functions. Numerical experiments indicate that selecting the upper bound within the range of $W = 500$ and $W = 800$ minimally impacts the validity and performance of the model identified by OPAL-surrogate, indicating the effectiveness of the

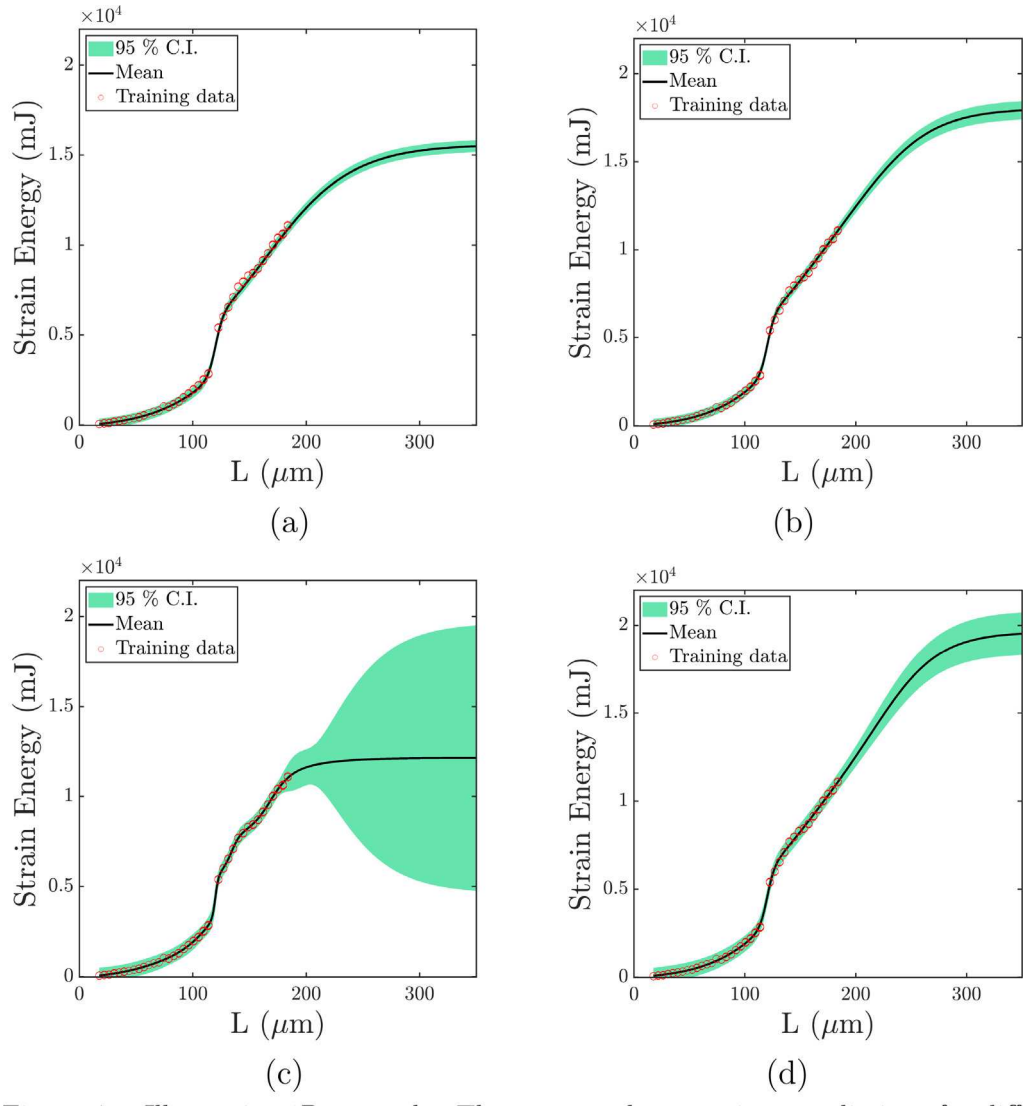

Fig. 4. Illustrative 1D example: The mean and uncertainty predictions for different fully connected BayesNN models with *Tanh* activation functions in Fig. 3(a) compared to the training data: (a) $D = 3$, $W = 720$, $\ln \pi_{evid}(\boldsymbol{D}|\boldsymbol{\xi}) = -3354$ (highest posterior model plausibility); (b) $D = 3$, $W = 800$, $\ln \pi_{evid}(\boldsymbol{D}|\boldsymbol{\xi}) = -3468$; (c) $D = 2$, $W = 1450$, $\ln \pi_{evid}(\boldsymbol{D}|\boldsymbol{\xi}) = -3877$; (d) $D = 1$, $W = 1840$, $\ln \pi_{evid}(\boldsymbol{D}|\boldsymbol{\xi}) = -3926$.

Table 1
Elasticity problem: Identifying the layerwise activation functions based on log-evidence $\pi_{evid}(\boldsymbol{D}|\ \boldsymbol{\xi}^l_{(\cdot)}, \boldsymbol{D}_c, \hat{\mathcal{M}}^l, \mathcal{S}, \mathcal{S}_c)$. All models are fully connected networks $M^l_{F(D,W=600)}$ with depth D and width $W = 600$.

Occam categories	BayesNN model	Log-evidence			
		ReLU	Leaky ReLU	Sigmoid	Tanh
$l = 1$	$D = 1$	−15732	−15045	−15317	**−14846**
	$D = 2$	−15320	−15246	−15314	**−14924**
$l = 2$	$D = 3$	−15835	**−14834**	−15427	−15340
	$D = 4$	−15723	−15256	−15417	**−15012**

proposed incremental search for initializing OPAL-surrogate. Accordingly, we categorize the BayesNN models with $W = [1, 600]$ based on the number of layers as a measure of model complexity, such that Category 1 encompassing networks with $D = \{1, 2\}$, Category 2 including those with $D = \{3, 4\}$, and so forth.

Occam category 1. Table 1 shows the values of model evidence $\pi_{evid}(\boldsymbol{D}|\ \boldsymbol{\xi}^l_{(\cdot)}, \boldsymbol{D}_c, \hat{\mathcal{M}}^l, \mathcal{S}, \mathcal{S}_c)$ for fully connected networks $M^l_{F(D,W=600)}$ with different activation functions. Accordingly, the *Tanh* function is selected for the first and second layers. Upon applying

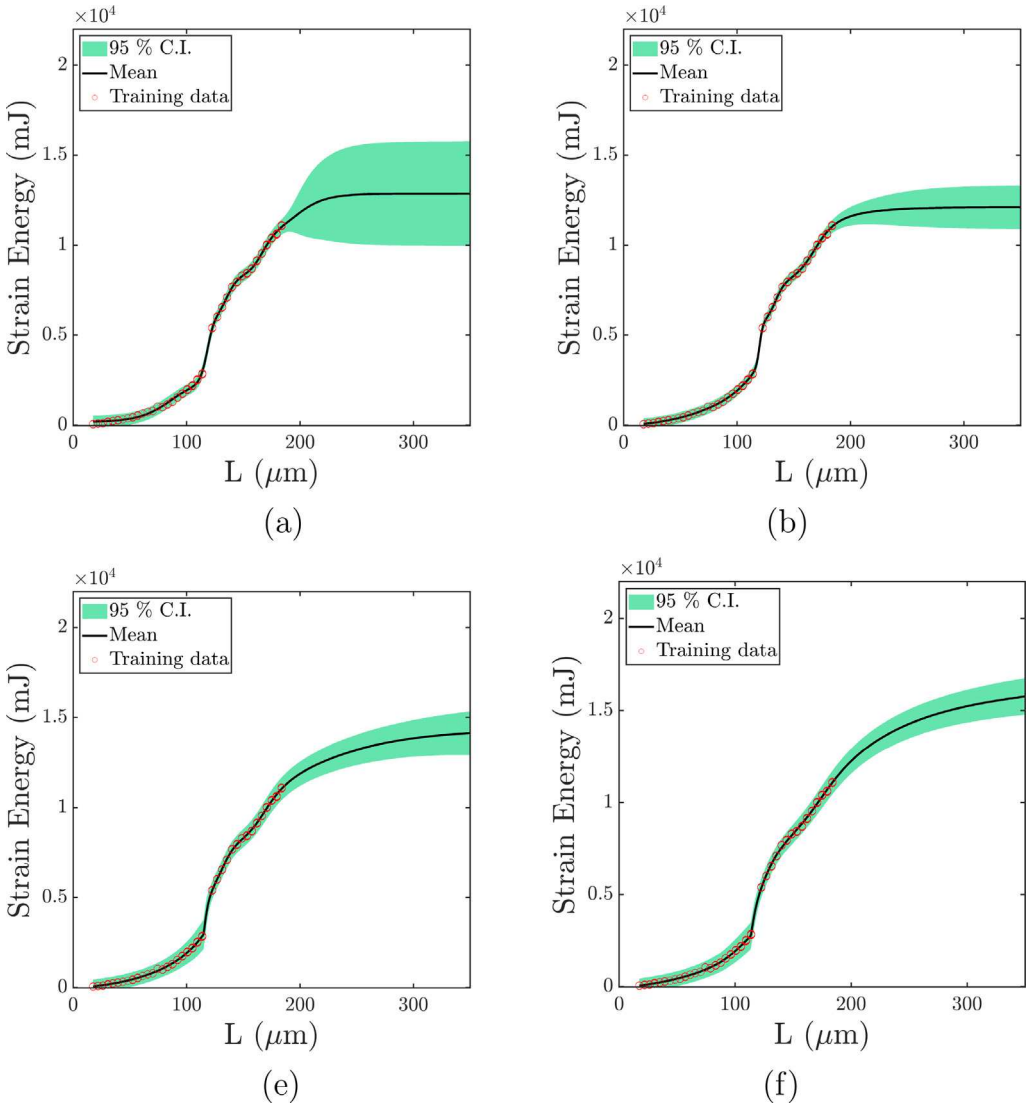

Fig. 5. Illustrative 1D example: Network sparsification results for the BayesNN models. The mean and uncertainty predictions for (a) the fully connected network with $D = 4$, $W = 1600$, and $\ln \pi_{evid}(\boldsymbol{D}|\boldsymbol{\xi}) = -3821$, and (b) the corresponding sparsified network using $TOL_{\theta} = 0.025$, resulting in the elimination of 22% of the parameters and yielding $\ln \pi_{evid}(\boldsymbol{D}|\boldsymbol{\xi}) = -3610$. (c) The fully connected network with $D = 6$, $W = 1700$, and $\ln \pi_{evid}(\boldsymbol{D}|\boldsymbol{\xi}) = -3874$, and (d) the corresponding sparsified network using $TOL_{\theta} = 0.05$, resulting in the elimination of 32% of the parameters and yielding $\ln \pi_{evid}(\boldsymbol{D}|\boldsymbol{\xi}) = -3545$.

sparsification to $M^1_{F(D=1,W=600)}$, the plausible model in this category M^1_I is identified that consists of 564 connections. According to Table 2, the sparsification results in elimination of 6% of the parameters and 0.6% improvement in log-evidence in M^1_I compared to $M^1_{F(D=1,W=600)}$. Fig. 7 illustrates the surrogate model predictions of both the fully connected and sparsified networks in comparison to the pre-training and training datasets. The observed increase in model prediction uncertainty at lower E_s is attributed to higher uncertainty in high-fidelity simulation data at smaller elastic modulus values captured by the surrogate model.

Next, we assess the credibility of M^1_I through a leave-out validation test. The comparison of validation measures for each leave-out data point with the corresponding validation tolerances reveals that M^1_I fails the validation test and is deemed an invalid model,

$L_{LO} = 179.4$ μm:

$$\mathrm{d}_{DKL} = 0.0112 \not\leqslant TOL_{DKL} = 0.008, \qquad \mathrm{d}_{CDF} = 52.31 \not\leqslant TOL_{CDF} = 45,$$

$L_{LO} = 183.8$ μm:

$$\mathrm{d}_{DKL} = 0.0108 \not\leqslant TOL_{DKL} = 0.008, \qquad \mathrm{d}_{CDF} = 60.52 \not\leqslant TOL_{CDF} = 45.$$

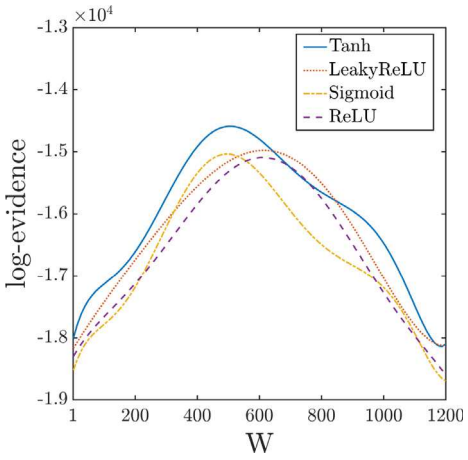

Fig. 6. Elasticity problem: The log-evidence (corresponding to model plausibility ρ) for BayesNN models with single layer ($D = 1$) with different widths and activation functions.

Table 2
Elasticity problem: Results from the implementation of OPAL-surrogate for the identification of the predictive BayesNN surrogate models. $M^l_{F(D,W)}$ denotes fully connected networks, and M^l_I represents the plausible sparsified model in each Occam Category l. The values of log-evidence $\pi_{evid}(\boldsymbol{D}|\boldsymbol{\xi}^l_{(\cdot)}, \boldsymbol{D}_c, \hat{\mathcal{M}}^l, S, S_c)$ and validation metrics corresponding to leave-out sets $L_{LO} = \{179.4, 183.8\}$ μm are presented. OPAL-surrogate identifies M^2_I at Category 2 as the "best" BayesNN surrogate model.

Occam	BayesNN	Activation	$M^l_{F(D,W)}$	M^l_I	d_{DKL}		d_{CDF}	
categories	models	function	log-evid	log-evid	179.4 μm	183.8 μm	179.4 μm	183.8 μm
$l = 1$	$D = 1$	Tanh	−14846	−14758	0.0112	0.0108	52.31	60.52
	$D = 2$	Tanh	−14924					
$l = 2$	$D = 3$	Leaky ReLU	−14834	−14691	**0.0073**	**0.0064**	**43.62**	**36.71**
	$D = 4$	Tanh	−15012					
$l = 3$	$D = 5$	Tanh	−15036	−14882	0.0124	0.0116	85.26	77.42
	$D = 6$	Tanh	−15133					
$l = 4$	$D = 7$	LeakyReLU	−15187					
	$D = 8$	Tanh	−15046	−14962	0.0287	0.0258	153.48	143.21

Occam category 2. Following the same procedure leads to the selection of the *Leaky ReLU* activation function for layer 3 and the *Tanh* function for layer 4. Subsequent sparsification of $M^2_{F(D=3,W=600)}$ results in the plausible model within this category, M^2_I, comprising $D = 3$ and 619 245 connections (approximately 14% parameters reduction upon sparsification). Fig. 8 presents the predictions of this surrogate model in comparison to high-fidelity datasets. Importantly, M^2_I successfully passes the validation test based on both validation metrics,

$L_{LO} = 179.4$ μm:

$$\mathrm{d}_{DKL} = 0.0069 \leq TOL_{DKL} = 0.008, \qquad \mathrm{d}_{CDF} = 40.42 \leq TOL_{CDF} = 45,$$

$L_{LO} = 183.8$ μm:

$$\mathrm{d}_{DKL} = 0.0061 \leq TOL_{DKL} = 0.008, \qquad \mathrm{d}_{CDF} = 34.55 \leq TOL_{CDF} = 45,$$

establishing it as the "best" predictive surrogate model for the given pre-training and training data sets. As depicted in Table 2, model M^2_I demonstrates significantly improved predictive performance compared to M^1_I, as evident from the validation metrics (a 37% reduction in d_{DKL} and a 30% decrease in d_{CDF}), and visually shown in Fig. 9.

Exploring higher categories. Although OPAL-surrogate concludes upon identifying the model M^2_I in Category 2, as depicted in Table 2 and Fig. 10, an investigation into the performance of more complex models in higher categories reveals compromised performance. Specifically, models M^3_I and M^4_I are deemed invalid based on the specified accuracy tolerance. This highlights the observation that larger networks with increased free parameters do not necessarily translate to improved predictive capabilities.

Prediction and forward UQ. Finally, we utilize the surrogate model, M^2_I, to predict and quantify uncertainty in the target QoI (strain energy) in the prediction scenario S_P. In Fig. 11(a), the surrogate model prediction at $E_s = 110$ MPa and domain size $L = 297.5$ μm is compared with one realization of the microstructure in S_P, as modeled by high-fidelity simulation. Fig. 11(b) presents the probability distributions of the QoI for uncertain elastic modulus of the silica aerogel $E_s \sim \mathcal{N}(107.5, 112.5)$, as obtained from previous studies [37]. The estimated mean and 95% CI of the QoI for the extrapolation prediction by the BayesNN surrogate

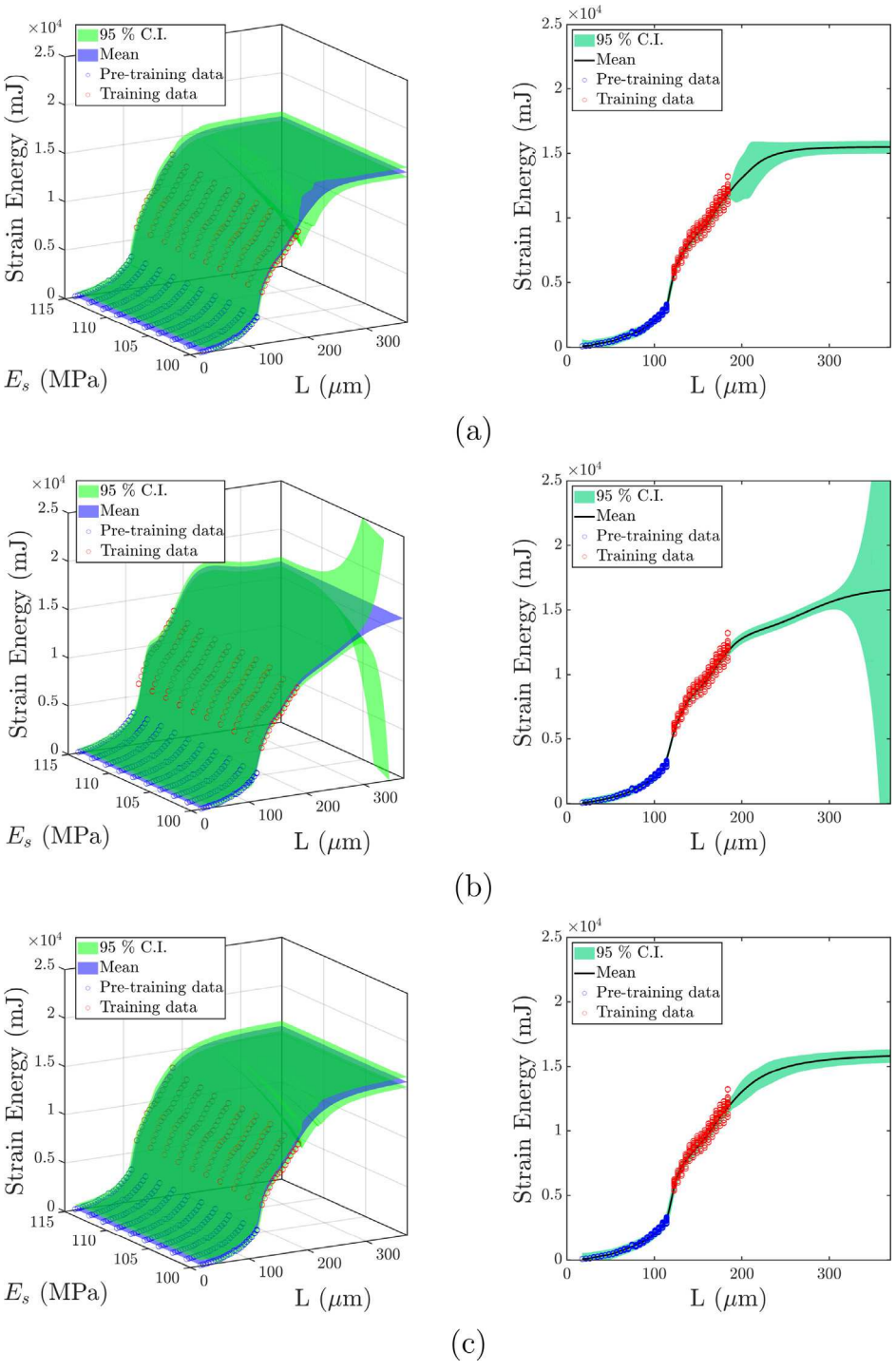

Fig. 7. Elasticity problem: The mean and uncertainty predictions of surrogate models in Category 1 for (a) fully-connected network $M^1_{F(D=1,W=600)}$; (b) fully-connected network $M^1_{F(D=2,W=600)}$; (c) plausible sparsified network M^1_I with $D = 1$ and 564 connections. The left panels show 3D plots, while the right panels depict the marginalized plots over the E_s.

model are $\mathbb{E}(QoI) = 16615, \quad CI(QoI) = [15565, 17665]$ mJ. However, considering only the mean of the surrogate model prediction (corresponding to deterministic neural network training), the estimated uncertainty becomes $\mathbb{E}(QoI) = 16660, \quad CI(QoI) = [16350, 16970]$ mJ. These results suggest that while the mean prediction of the surrogate model demonstrates relatively acceptable performance, an evaluation of model prediction uncertainty using BayesNN highlights the intrinsic limitation of neural networks in extrapolation.

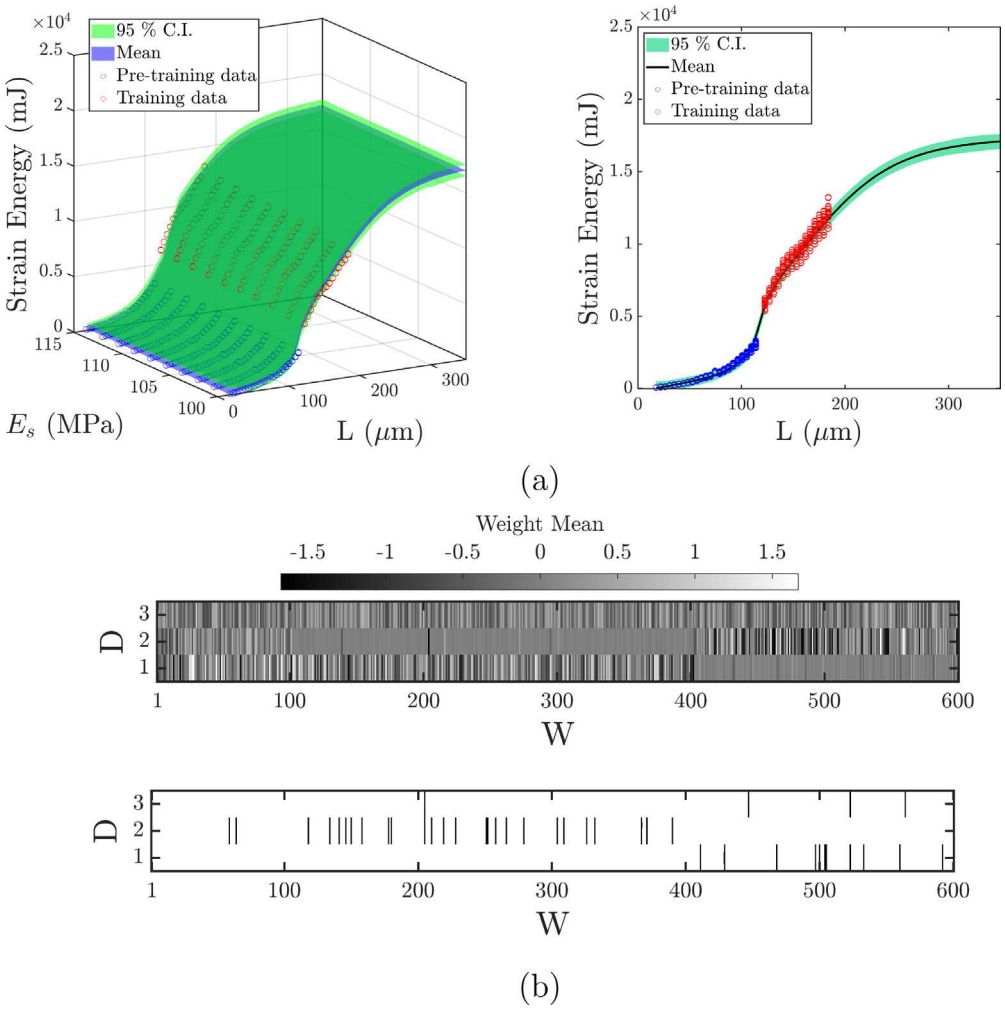

Fig. 8. Elasticity problem: (a) The mean and uncertainty predictions of the plausible sparsified network M_I^2 in Category 2, identified as the "best" surrogate model by OPAL-surrogate. The network consists of 3 layers, 619 245 connections, and *Tanh*, *Tanh*, *Leaky ReLU* activation functions for layers 1 to 3, respectively. (b) Visualization of weights pattern for $M^2_{F(D=3,W=600)}$ across network depth D and width W in the top panel. The bottom panel illustrates the sparsified pattern of M_I^2, with removed connections marked in black from each hidden layer using $TOL_\theta = 0.05$.

5.2. Direct numerical simulations of turbulent combustion

Our second application focuses on the combustion dynamics of shear-induced ablation in solid fuels used in hybrid rocket motors. Specifically, we leverage the Ablative Boundary Layers at the Exascale (ABLATE) software framework [79], which integrates direct numerical simulations of turbulent reacting flows with thermochemical species equations. The focal point of our investigation centers on modeling slab burner experiments according to the setup in [84,85], designed to study the reacting boundary layer combustion in hybrid rockets and measuring fuel regression rates. Turbulent combustion involves complex interactions among chemical reactions, flame structures, radiation, and soot, occurring on scales much smaller than flow transport. ABLATE's Navier–Stokes solver comprehensively models these phenomena by simulating fully compressible and reactive gas phases.

Fig. 12 depicts a representative simulation result from a 2D slab burner employed in this study, with Polymethyl methacrylate serving as the solid fuel and O_2 as the oxidizer in ABLATE simulations. The key output of interest is the fuel regression rate, denoted as $\dot{r}$, representing the rate of fuel recession during combustion and a critical parameter influencing motor thrust and hybrid rocket motor geometry. Given the high computational costs associated with ABLATE simulations, the ultimate aim of the surrogate model is to facilitate Bayesian calibration and validation of ABLATE by utilizing regression rate measurements from slab burner experiments. To this end, we develop a BayesNN surrogate for the fuel regression rate, incorporating two input parameters of ABLATE. The first input is the oxidizer flux $G = [5, 20]$ kg/m^2 s, corresponding to the range of inlet velocities in the slab burner experiment [84,85], and the second input parameter is the latent heat of vaporization $l_v = [6 \times 10^5, 11 \times 10^5]$ J/kg. The training dataset $\boldsymbol{D}$, comprising $N_D = 64$ simulation ensembles, captures the fuel regression rate $\dot{r}$ at a critical location on the slab boundary (illustrated in Fig. 12),

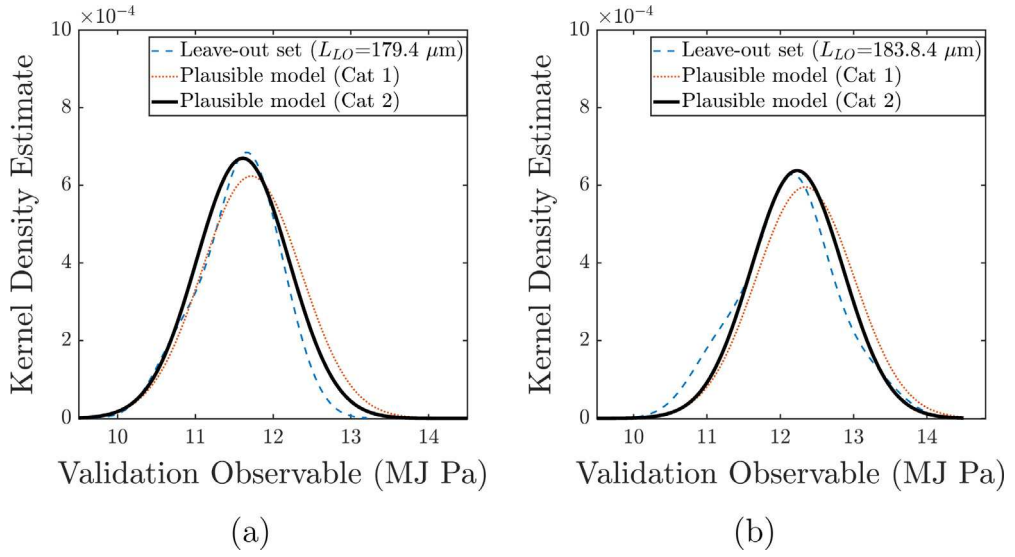

Fig. 9. Elasticity problem: Comparison of the observables $\mathcal{Z}_D$ in (36) corresponding to leave-out validation sets and the predicted $\mathcal{Z}_M$ from plausible models M_I^1 and M_I^2 for (a) $L_{LO} = 179.4$ μm; (b) $L_{LO} = 183.8$ μm.

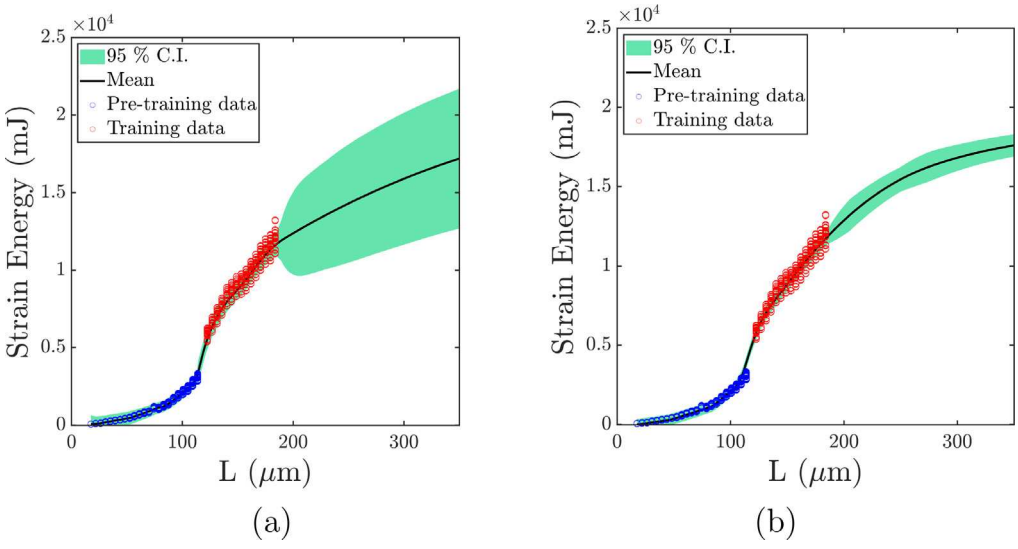

Fig. 10. Elasticity problem: The mean and uncertainty predictions, marginalized over the E_s, of surrogate models in Categories 3 and 4: (a) plausible sparsified network M_I^3 with $D = 5$ and 1 264 240 connections; (b) fully-connected network $M^4_{F(D=7,W=600)}$. Both panels depict the marginalized plots over the E_s.

with the inputs G and l_v sampled using Latin hypercube sampling. For credibility assessment, we consider the leave-out set $\boldsymbol{D}_{LO}$ containing data points within $G_{LO} = [9.89, 14.63]$ kg/m^2 s. The validation observables are,

$$\mathcal{Z}_D = \int_{G_{LO}^{low}}^{G_{LO}^{up}} \int_{l_v^{low}}^{l_v^{up}} u_{\boldsymbol{D}}\, dl_v\, dG, \qquad \mathcal{Z}_M = \int_{G_{LO}^{low}}^{G_{LO}^{up}} \int_{l_v^{low}}^{l_v^{up}} u_{\boldsymbol{\theta}}\, dl_v\, dG, \tag{39}$$

where $u_{\boldsymbol{D}}$ and $u_{\boldsymbol{\theta}}$ are the fuel regression rate $\dot{r}$, evaluated using the high-fidelity simulations and surrogate models, respectively, and $G_{LO}^{low} = 9.89$ kg/m^2 s, $G_{LO}^{up} = 14.63$ kg/m^2 s, $l_v^{low} = 6 \times 10^5$ J/kg, and $l_v^{up} = 11 \times 10^5$ J/kg.

OPAL-surrogate demonstration. Following the approach outlined in Section 5.1.2, we set the upper limit for the width at $W = 300$ in the initial BayesNN surrogate model set $\mathcal{M}$, guided by Fig. 13. Given the smaller dataset, a more strict categorization strategy is employed for models with $W = [1, 300]$, with each category corresponding to a distinct width range.

Table 3 presents the results of OPAL-surrogate for the first three categories. Figs. 14 and 15(a) illustrate the predictions of both fully connected and sparsified networks in Category 1, comparing them to the training datasets. The plausible model M_I^1 comprises 245 connections, approximately 19% parameter reduction from $M^1_{F(D=1,W=300)}$ following sparsification with $TOL_{\boldsymbol{\theta}} = 0.075$. However,

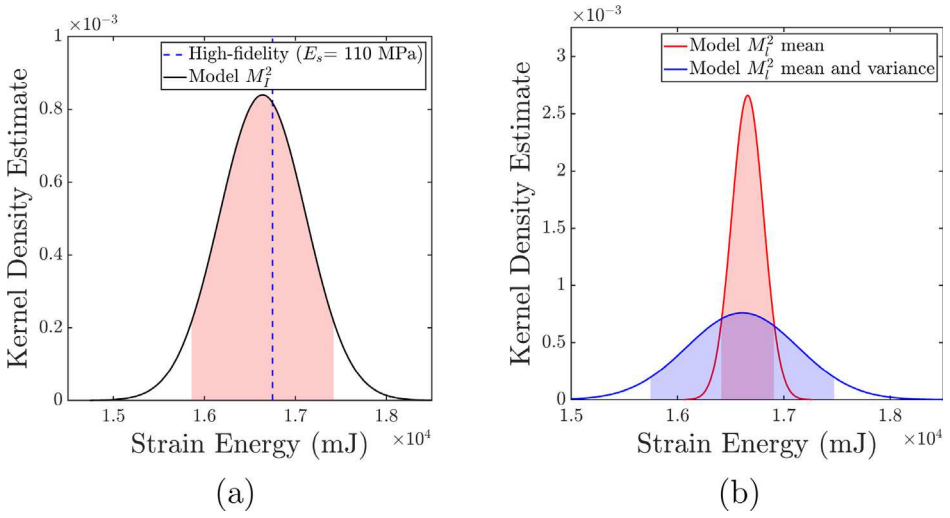

Fig. 11. Elasticity problem: Prediction of the QoI, strain energy, in the prediction scenario S_P, utilizing the BayesNN surrogate model M_I^2, considering: (a) elastic modulus $E_s = 110$ MPa and in comparison with high-fidelity simulation for one microstructure realization; (b) uncertain elastic modulus $E_s \sim \mathcal{N}(107.5, 112.5)$ MPa and with and without the inclusion of uncertainty bounds in the model predictions.

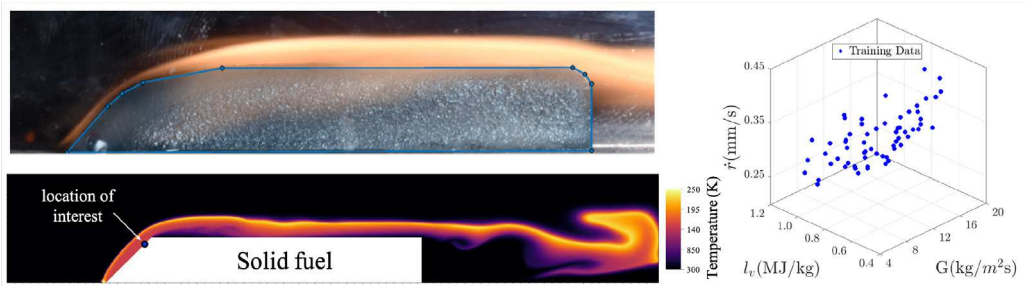

Fig. 12. Turbulent combustion problem: The top-left panel illustrates the side profile of the solid fuel in the 2D slab burner experiment [84,85]. The bottom-left panel displays a representative 2D ABLATE simulation result of the slab burner, highlighting the location on the fuel surface for obtaining fuel regression rate ($\dot{r}$) training data. The right panel shows the training data used for surrogate modeling, comprising an ensemble of $\dot{r}$ for various values of the oxidizer flux (G) and latent heat of vaporization (l_v).

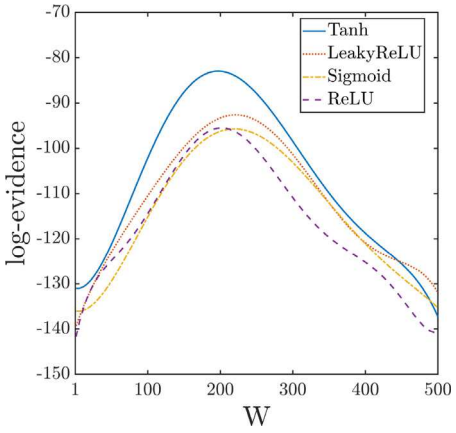

Fig. 13. Turbulent combustion problem: The log-evidence for BayesNN models with a single layer ($D = 1$) with different widths and activation functions.

considering $TOL_{DKL} = 0.01$, this model is deemed invalid. Fig. 15(a) illustrates the predictions of M_I^1 compared to the leave-out data for credibility assessment.

Table 3
Turbulent combustion problem: Results of OPAL-surrogate implementation for the identification of M_I^2 at Category 2 as the "best" BayesNN surrogate model.

Occam categories	BayesNN models	Activation function	$M^l_{F(D,W)}$ log-evid	M^l_I log-evid	d_{DKL}
$l = 1$	$D = 1$	Tanh	−99.5	−95.5	0.0127
$l = 2$	$D = 2$	Tanh	−94.7	−90.6	**0.0084**
$l = 3$	$D = 3$	LeakyReLU	−103.6	−98.7	0.0156

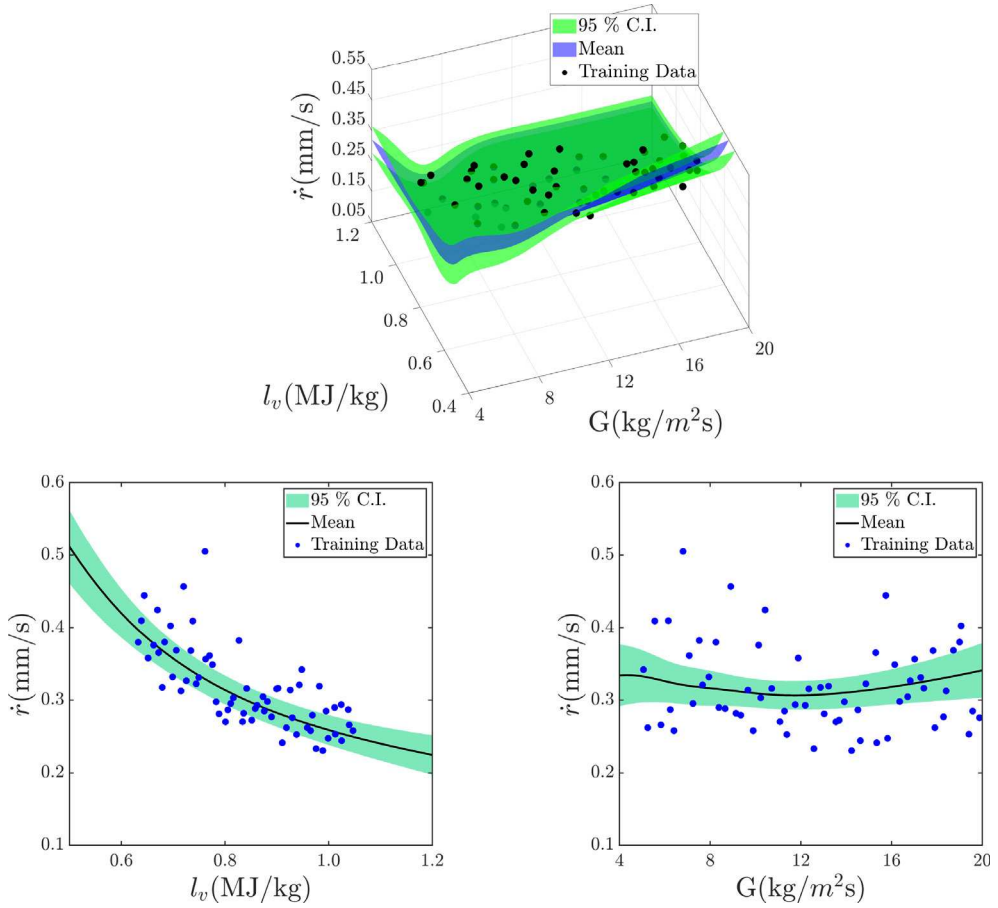

Fig. 14. Turbulent combustion problem: The mean and uncertainty predictions of the fully-connected network $M^1_{F(D=1,W=300)}$ in Category 1. The top panel shows the 3D plot and the bottom panels display the marginalized plots over the l_v and G.

Moving to Category 2, the *Tanh* activation function is chosen for the second layer. Sparsification of $M^2_{F(D=2,W=300)}$ with 90 000 connections, results in the plausible model M_I^2 for this category, comprising $D = 2$ and 65 025 connections, representing approximately 28% reduction in parameters after sparsification with $TOL_\theta = 0.12$. As shown in Table 3, M_I^2 successfully passes the validation test establishing it as the "best" predictive surrogate model. Fig. 16 illustrates the predictions of M_I^2 compared to the training data and Fig. 17 showes this model's predictive perfroamnce compared to the leave-out dataset. The means of the posteriors of the inference hyperparameters shown in this figure are obtained by maximizing the evidence in (24), and the posterior variances are approximated using (25). As indicated in Table 3, model M_I^2 demonstrates significantly improved predictive performance compared to M_I^1 and M_I^3 (see Fig. 19) in both lower and higher categories. The comparison of validation observables is presented in Fig. 18 (see Fig. 17).

6. Conclusions

This study introduces OPAL-surrogate, a systematic framework for identifying predictive BayesNN surrogate models within the expansive space of potential models characterized by diverse architectures and hyperparameters. Leveraging hierarchical Bayesian inferences and the concept of model plausibility, OPAL-surrogate efficiently and adaptively adjusts model complexity until satisfying

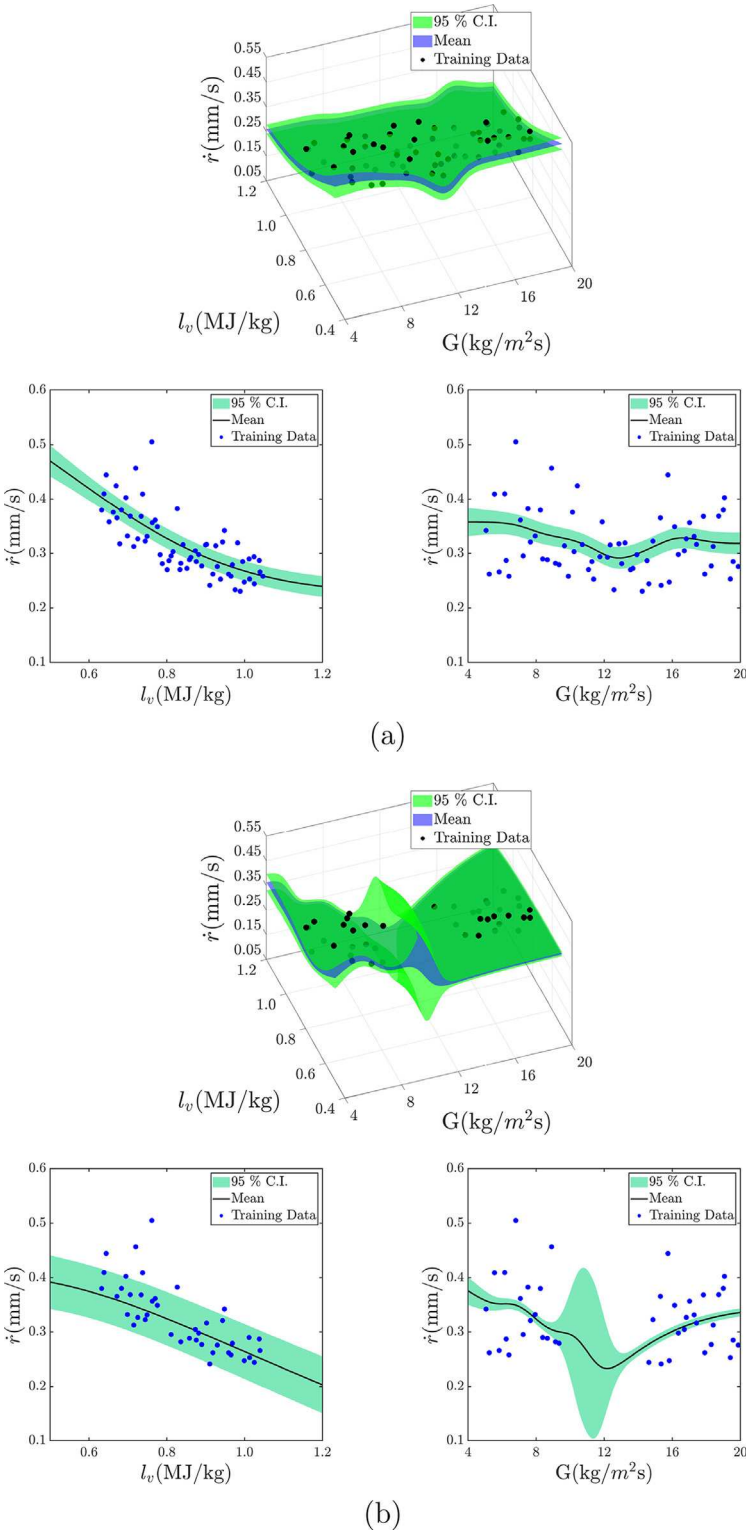

Fig. 15. Turbulent combustion problem: The mean and uncertainty predictions of the plausible sparsified network M_l^1 in Category 1 with $D = 1$ and 245 connections in comparison with (a) training data set; (b) leave-out data set. The top panel shows the 3D plot and the bottom panels display the marginalized plots over the l_v and G.

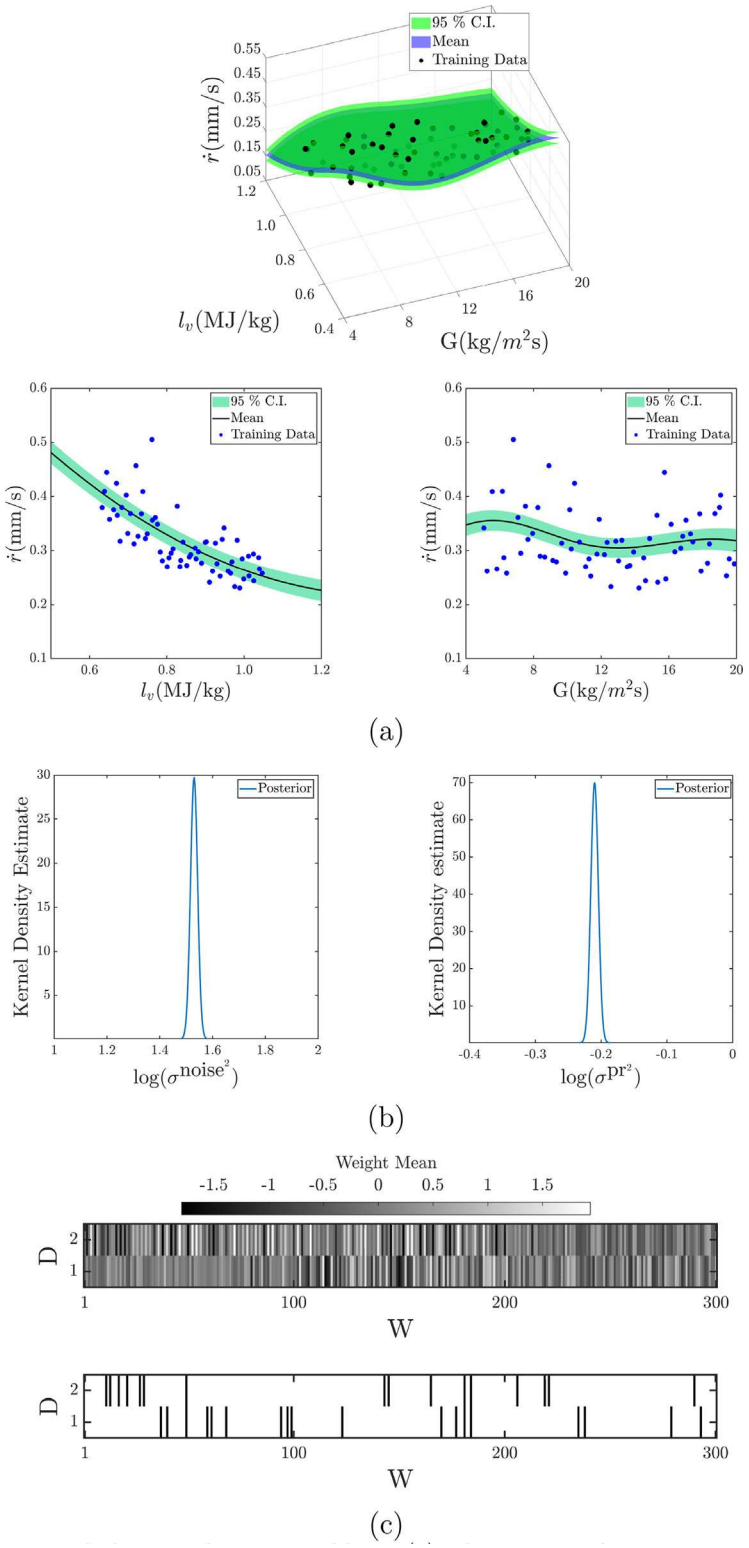

Fig. 16. Turbulent combustion problem: (a) The mean and uncertainty predictions of the plausible sparsified network M_I^2 in Category 2, identified as the "best" surrogate model by OPAL-surrogate. The network consists of 2 layers, 65 025 connections, and *Tanh* activation functions for both layers. (b) Posterior of the inference hyper-parameters considering the priors $\pi_{pr}(\log(\sigma^{\text{pr}^2})) = \mathcal{U}(-10, 10)$ and $\pi_{pr}(\log(\sigma^{\text{noise}^2})) = \mathcal{U}(-6, 6)$. (c) Visualization of weights pattern for $M^2_{F(D=2,W=300)}$ across network depth D and width W in the top panel. The bottom panel illustrates the sparsified pattern of M_I^2, with removed connections marked in black from each hidden layer using $TOL_\theta = 0.12$.

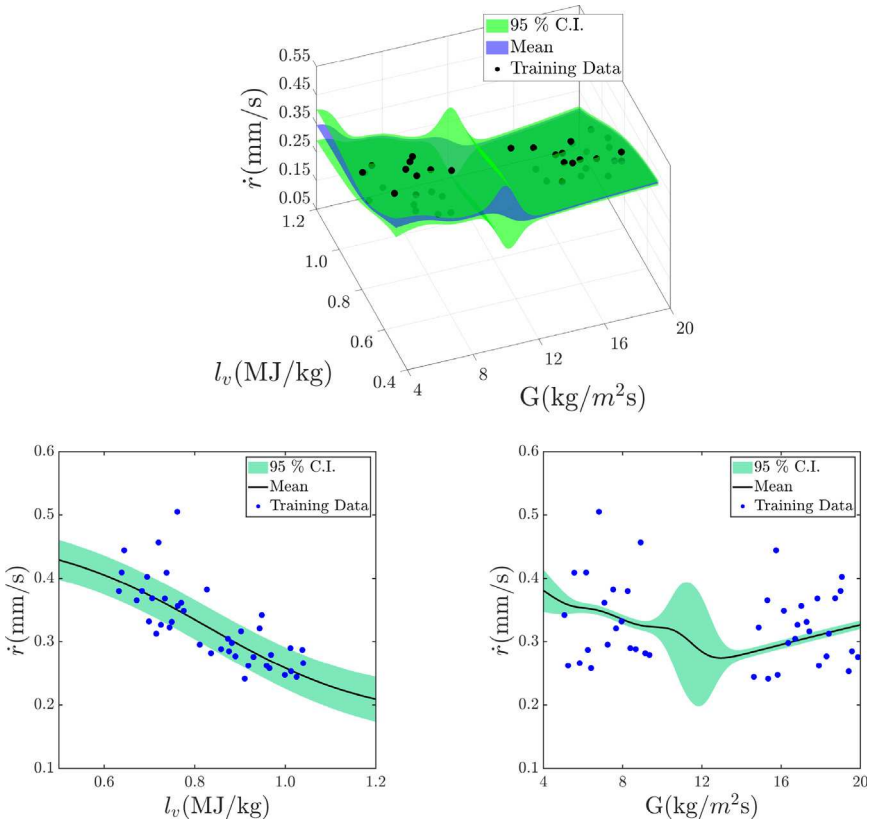

Fig. 17. Turbulent combustion problem: The mean and uncertainty predictions of the plausible sparsified network M_I^2 in Category 2 in comparison with the leave-out data. The top panel shows the 3D plot and the bottom panels display the marginalized plots over the l_v and G.

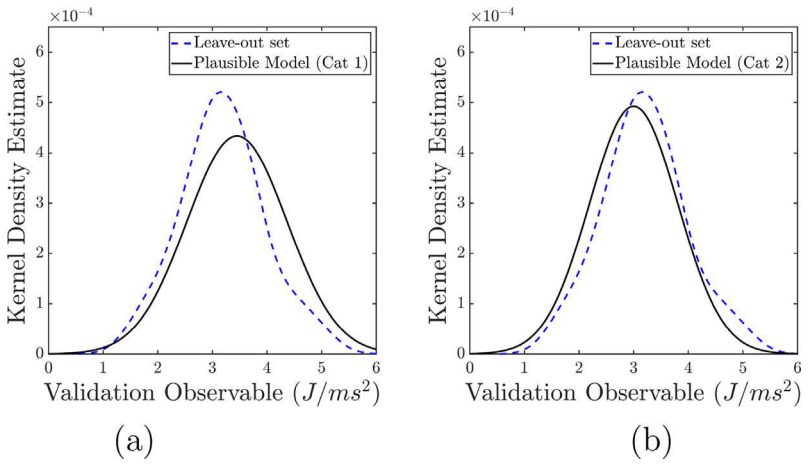

Fig. 18. Turbulent combustion problem: Comparison of the observables $\mathcal{Z}_D$ in (39) corresponding to leave-out validation sets and the predicted $\mathcal{Z}_M$ from plausible models M_I^1 and M_I^2.

validation criteria. We also stress the significance of well-organized neural network parameters, empowering the surrogate model to effectively capture multiscale interactions encoded in the training data for physics-based simulations. To achieve this, we propose a method involving the sequential addition of fully connected layers with large widths and the elimination of irrelevant weights through an effective network sparsification guided by model evidence.

Two applications of OPAL-surrogate demonstrate that the identified architecture and hyperparameters of the BayesNN model, achieved through a balanced trade-off between model complexity and validity, result in enhanced accuracy and reliability in

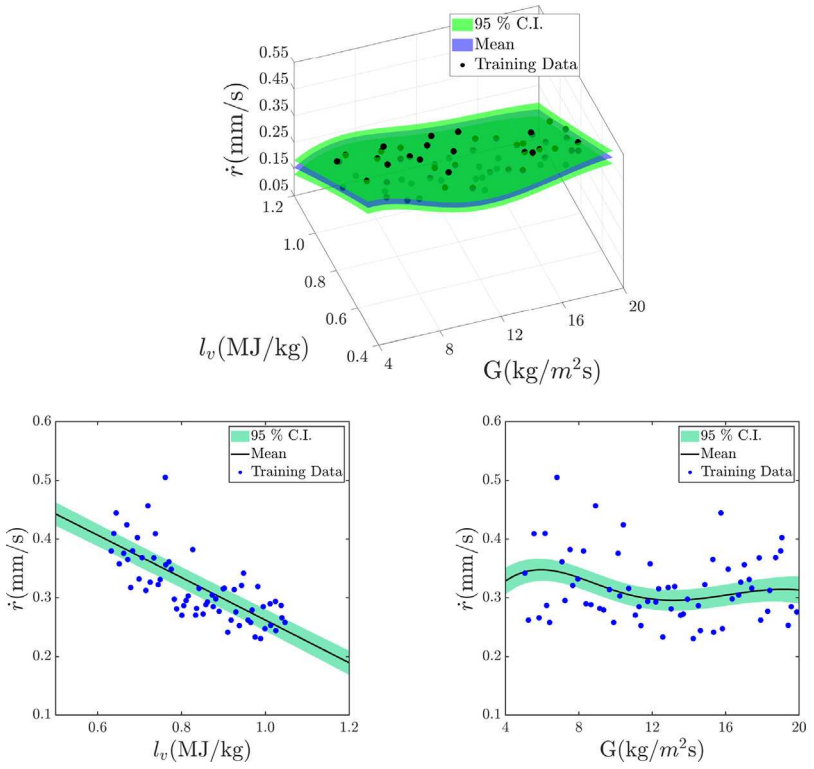

Fig. 19. Turbulent combustion problem: The mean and uncertainty predictions of the plausible sparsified network M_1^3 in Category 3 with $D = 3$ and 135 000 connections in comparison with training data. The top panel shows the 3D plot and the bottom panels display the marginalized plots over the l_v and G.

predictions. The first example involves surrogate modeling of elastic deformation in porous materials, aiming to facilitate quantifying uncertainty in unobservable QoI for domain sizes where high-fidelity simulation is computationally prohibitive. The results suggest that, despite a substantial training dataset comprising 1500 data points and constructing the prior for network parameters from the pre-training dataset, the extrapolation prediction capability of the identified BayesNN model remains credible only up to 1.5 times the size of the training domain. Beyond this point, the accuracy and reliability of the prediction rapidly decline, highlighting the necessity of imposing physical constraints to enhance the extrapolation abilities of neural networks, e.g., [86–88]. The second numerical experiment involves a turbulent combustion flow model of a slab burner, where OPAL-surrogate successfully identifies the BayesNN surrogate model for the fuel regression rate based on 64 training data points.

We highlight a fundamental misconception contributing to overconfident predictions — the notion that the complexity of a predictive model must always be limited when training data is scarce. In contrast, the OPAL-surrogate framework relies on Bayesian inference without the constraint of modifying the model and prior based on the data volume. In fact, we argue that there is never enough high-fidelity data from physical simulations for surrogate model construction. Thus, OPAL-surrogate embodies an open-ended inference approach, continuously refining the predictive surrogate model, as that additional data and information could potentially *falsify* a model initially presumed to be valid. In constructing the initial set of possible models, [26,89,90] propose incorporating models we genuinely believe in, along with every conceivable sub-model, ensuring that the model selection strategy can identify the sub-model that best explains the data.

In advancing the surrogate modeling using the proposed framework in this work, several avenues can be explored in future studies. Firstly, more accurate Bayesian solutions may be adopted, recognizing that the posterior distribution is only asymptotically normally distributed as the number of data points approaches infinity, and in neural networks, they might be multi-modal. Despite this, given the Laplace approximation's high efficiency and scalability compared to other existing algorithms, it may be worthwhile to expand the definition of the BayesNN model, incorporating various solutions with differing complexities into the initial model set. Subsequently, OPAL-surrogate can discern the one that provides valid predictions. This approach is justified by the primary goal of the surrogate model, which is to faithfully approximate the predictive distribution rather than the accuracy of parameter inference. Secondly, there is a need for comprehensive investigations into effective methods for revealing neural network sparsity patterns associated with multiscale physical phenomena. This can involve exploring techniques like probabilistic sparse masks, e.g., [74,91] and leveraging architecture search algorithms, e.g., [34], with the potential to enhance sparsification efficiency after replacing the performance measure with model evidence. Lastly, in the future, we aim to enhance the versatility of OPAL-surrogate to accommodate a broader range of possible models, e.g., different classes of neural operators [92–97]. This expansion could significantly reinforce its capacity to identify the appropriate surrogate model for a given problem.

In conclusion, this study highlights the substantial challenges associated with the discovery and assessing the credibility of neural network-based surrogate models for complex multiscale and multiphysics simulations. The introduced framework aims to tackle some of these challenges by highlighting the crucial interplay between model complexity and rigorous validation. It provides a foundation for future research to enhance its efficacy and extend its applicability across diverse classes of surrogate models.

CRediT authorship contribution statement

Pratyush Kumar Singh: Writing – review & editing, Visualization, Validation, Software, Investigation, Formal analysis, Data curation. **Kathryn A. Farrell-Maupin:** Conceptualization. **Danial Faghihi:** Writing – review & editing, Writing – original draft, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Danial Faghihi and Pratyush Kumar Singh report financial support was provided by National Science Foundation (NSF). Danial Faghihi also acknowledges the partial support from the Department of Energy's National Nuclear Security Administration (NNSA)

Data availability

Data will be made available on request.

Acknowledgments

DF and PKS extend their sincere gratitude for the financial support received from the U.S. National Science Foundation (NSF) CAREER Award CMMI-2143662. DF also appreciates the partial support from the U.S. Department of Energy's (DoE) National Nuclear Security Administration (NNSA), USA under the Predictive Science Academic Alliance Program III (PSAAP III) DE-NA0003961. Additionally, the authors would like to acknowledge the support provided by the Center for Computational Research at the University at Buffalo.

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525. This paper describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the U.S. Department of Energy or the United States Government.

References

[1] R.K. Tripathy, I. Bilionis, Deep UQ: Learning deep neural network surrogate models for high dimensional uncertainty quantification, J. Comput. Phys. 375 (2018) 565–588.
[2] Y. Zhu, N. Zabaras, Bayesian deep convolutional encoder–decoder networks for surrogate modeling and uncertainty quantification, J. Comput. Phys. 366 (2018) 415–447.
[3] G. Georgalis, K. Retfalvi, P.E. Desjardin, A. Patra, Combined data and deep learning model uncertainties: An application to the measurement of solid fuel regression rate, Int. J. Uncertain. Quantif. 13 (5) (2023) http://dx.doi.org/10.1615/int.j.uncertaintyquantification.2023046610.
[4] L. Scarabosio, B. Wohlmuth, J.T. Oden, D. Faghihi, Goal-oriented adaptive modeling of random heterogeneous media and model-based multilevel Monte Carlo methods, Comput. Math. Appl. 78 (8) (2019) 2700–2718.
[5] Y. Li, Y. Wang, L. Yan, Surrogate modeling for Bayesian inverse problems based on physics-informed neural networks, J. Comput. Phys. 475 (2023) 111841.
[6] L. Cao, K. Wu, J.T. Oden, P. Chen, O. Ghattas, Bayesian model calibration for diblock copolymer thin film self-assembly using power spectrum of microscopy data and machine learning surrogate, Comput. Methods Appl. Mech. Engrg. 417 (2023) 116349.
[7] D. Luo, T. O'Leary-Roseberry, P. Chen, O. Ghattas, Efficient PDE-constrained optimization under high-dimensional uncertainty using derivative-informed neural operators, 2023, arXiv preprint arXiv:2305.20053.
[8] J. Tan, D. Faghihi, A scalable framework for multi-objective PDE-constrained design of building insulation under uncertainty, Comput. Methods Appl. Mech. Engrg. 419 (2024) 116628.
[9] A. Chattopadhyay, M. Gray, T. Wu, A.B. Lowe, R. He, OceanNet: A principled neural operator-based digital twin for regional oceans, 2023, arXiv preprint arXiv:2310.00813.
[10] Q. He, M. Perego, A.A. Howard, G.E. Karniadakis, P. Stinis, A hybrid deep neural operator/finite element method for ice-sheet modeling, 2023, arXiv preprint arXiv:2301.11402.
[11] M.G. Kapteyn, J.V. Pretorius, K.E. Willcox, A probabilistic graphical model foundation for enabling predictive digital twins at scale, Nat. Comput. Sci. 1 (5) (2021) 337–347.
[12] S. Mowlavi, S. Nabi, Optimal control of PDEs using physics-informed neural networks, J. Comput. Phys. 473 (2023) 111731.
[13] T. Zohdi, A digital-twin and machine-learning framework for precise heat and energy management of data-centers, Comput. Mech. 69 (6) (2022) 1501–1516.
[14] K. Wu, T. O'Leary-Roseberry, P. Chen, O. Ghattas, Large-scale Bayesian optimal experimental design with derivative-informed projected neural network, J. Sci. Comput. 95 (1) (2023) 30.
[15] J. Stuckner, M. Piekenbrock, S.M. Arnold, T.M. Ricks, Optimal experimental design with fast neural network surrogate models, Comput. Mater. Sci. 200 (2021) 110747.
[16] A. Arzani, L. Yuan, P. Newell, B. Wang, Interpreting and generalizing deep learning in physics-based problems with functional linear models, 2023, arXiv preprint arXiv:2307.04569.

[17] L. Yuan, H.S. Park, E. Lejeune, Towards out of distribution generalization for problems in mechanics, Comput. Methods Appl. Mech. Engrg. 400 (2022) 115569.
[18] W. Samek, G. Montavon, S. Lapuschkin, C.J. Anders, K.-R. Müller, Explaining deep neural networks and beyond: A review of methods and applications, Proc. IEEE 109 (3) (2021) 247–278.
[19] X. Zhong, B. Gallagher, S. Liu, B. Kailkhura, A. Hiszpanski, T.Y.-J. Han, Explainable machine learning in materials science, NPJ Comput. Mater. 8 (1) (2022) 204.
[20] T. Oden, R. Moser, O. Ghattas, Computer predictions with quantified uncertainty, part I, SIAM News 43 (9) (2010).
[21] I. Babuška, F. Nobile, R. Tempone, A systematic approach to model validation based on Bayesian updates and prediction related rejection criteria, Comput. Methods Appl. Mech. Engrg. 197 (29) (2008) 2517–2539, http://dx.doi.org/10.1016/j.cma.2007.08.031, URL https://www.sciencedirect.com/science/article/pii/S0045782507005142. Validation Challenge Workshop.
[22] J.T. Oden, I. Babuška, D. Faghihi, Predictive computational science: Computer predictions in the presence of uncertainty, in: Encyclopedia of Computational Mechanics Second Edition, Wiley Online Library, 2017, pp. 1–26.
[23] J. Tan, B. Liang, P.K. Singh, K.A. Farrell-Maupin, D. Faghihi, Toward selecting optimal predictive multiscale models, Comput. Methods Appl. Mech. Engrg. 402 (2022) 115517.
[24] M. Yaseen, X. Wu, Quantification of deep neural network prediction uncertainties for VVUQ of machine learning models, Nucl. Sci. Eng. 197 (5) (2023) 947–966.
[25] J.M. Twomey, A.E. Smith, et al., Validation and verification, in: Artificial Neural Networks for Civil Engineers: Fundamentals and Applications, ASCE Press, New York, 1997, pp. 44–64.
[26] D.J. MacKay, Probable networks and plausible predictions-a review of practical Bayesian methods for supervised neural networks, Netw.: Comput. Neural Syst. 6 (3) (1995) 469.
[27] R.M. Neal, Bayesian Learning for Neural Networks, vol. 118, Springer Science & Business Media, 2012.
[28] T. Elsken, J.H. Metzen, F. Hutter, Neural architecture search: A survey, J. Mach. Learn. Res. 20 (1) (2019) 1997–2017.
[29] Y. Liu, Y. Sun, B. Xue, M. Zhang, G.G. Yen, K.C. Tan, A survey on evolutionary neural architecture search, IEEE Trans. Neural Netw. Learn. Syst. (2021).
[30] B. Wang, Y. Sun, B. Xue, M. Zhang, A hybrid differential evolution approach to designing deep convolutional neural networks for image classification, in: AI 2018: Advances in Artificial Intelligence: 31st Australasian Joint Conference, Wellington, New Zealand, December 11-14, 2018, Proceedings 31, Springer, 2018, pp. 237–250.
[31] A. Ghosh, N.D. Jana, S. Mallik, Z. Zhao, Designing optimal convolutional neural network architecture using differential evolution algorithm, Patterns 3 (9) (2022) 100567.
[32] H. Mendoza, A. Klein, M. Feurer, J.T. Springenberg, F. Hutter, Towards automatically-tuned neural networks, in: Workshop on Automatic Machine Learning, PMLR, 2016, pp. 58–65.
[33] F. Hutter, L. Kotthoff, J. Vanschoren (Eds.), Automated Machine Learning - Methods, Systems, Challenges, Springer, 2019.
[34] T. Akiba, S. Sano, T. Yanase, T. Ohta, M. Koyama, Optuna: A next-generation hyperparameter optimization framework, in: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2019.
[35] D. Faghihi, S. Sarkar, M. Naderi, J.E. Rankin, L. Hackel, N. Iyyer, A probabilistic design method for fatigue life of metallic component, ASCE-ASME J. Risk Uncertain. Eng. Syst. B 4 (3) (2018).
[36] J. Tan, U. Villa, N. Shamsaei, S. Shao, H.M. Zbib, D. Faghihi, A predictive discrete-continuum multiscale model of plasticity with quantified uncertainty, Int. J. Plast. 138 (2021) 102935.
[37] J. Tan, P. Maleki, L. An, M. Di Luigi, U. Villa, C. Zhou, S. Ren, D. Faghihi, A predictive multiphase model of silica aerogels for building envelope insulations, Comput. Mech. 69 (6) (2022) 1457–1479.
[38] J.T. Oden, E.E. Prudencio, A. Hawkins-Daarud, Selection and assessment of phenomenological models of tumor growth, Math. Models Methods Appl. Sci. 23 (07) (2013) 1309–1338.
[39] P.K. Jha, L. Cao, J.T. Oden, Bayesian-based predictions of COVID-19 evolution in Texas using multispecies mixture-theoretic continuum models, Comput. Mech. 66 (5) (2020) 1055–1068.
[40] B. Liang, J. Tan, L. Lozenski, D.A. Hormuth II, T.E. Yankeelov, U. Villa, D. Faghihi, Bayesian inference of tissue heterogeneity for individualized prediction of glioma growth, IEEE Trans. Med. Imaging (2023).
[41] E.A. Lima, D. Faghihi, R. Philley, J. Yang, J. Virostko, C.M. Phillips, T.E. Yankeelov, Bayesian calibration of a stochastic, multiscale agent-based model for predicting in vitro tumor growth, PLoS Comput. Biol. 17 (11) (2021) e1008845.
[42] E. Prudencio, P. Bauman, D. Faghihi, K. Ravi-Chandar, J. Oden, A computational framework for dynamic data-driven material damage control, based on Bayesian inference and model selection, Internat. J. Numer. Methods Engrg. 102 (3–4) (2015) 379–403.
[43] K. Farrell, J.T. Oden, D. Faghihi, A Bayesian framework for adaptive selection, calibration, and validation of coarse-grained models of atomistic systems, J. Comput. Phys. 295 (2015) 189–208.
[44] J.T. Oden, E.A. Lima, R.C. Almeida, Y. Feng, M.N. Rylander, D. Fuentes, D. Faghihi, M.M. Rahman, M. DeWitt, M. Gadde, et al., Toward predictive multiscale modeling of vascular tumor growth, Arch. Comput. Methods Eng. 23 (4) (2016) 735–779.
[45] D.J. MacKay, Bayesian interpolation, Neural Comput. 4 (3) (1992) 415–447.
[46] C.M. Bishop, Neural Networks for Pattern Recognition, Oxford University Press, 1995.
[47] A.F. Psaros, X. Meng, Z. Zou, L. Guo, G.E. Karniadakis, Uncertainty quantification in scientific machine learning: Methods, metrics, and comparisons, J. Comput. Phys. (2023) 111902.
[48] L. Yang, X. Meng, G.E. Karniadakis, B-PINNs: Bayesian physics-informed neural networks for forward and inverse PDE problems with noisy data, J. Comput. Phys. 425 (2021) 109913, http://dx.doi.org/10.1016/j.jcp.2020.109913, URL https://www.sciencedirect.com/science/article/pii/S0021999120306872.
[49] A. Olivier, M.D. Shields, L. Graham-Brady, Bayesian neural networks for uncertainty quantification in data-driven materials modeling, Comput. Methods Appl. Mech. Engrg. 386 (2021) 114079.
[50] K. Linka, A. Schäfer, X. Meng, Z. Zou, G.E. Karniadakis, E. Kuhl, Bayesian physics informed neural networks for real-world nonlinear dynamical systems, Comput. Methods Appl. Mech. Engrg. 402 (2022) 115346, http://dx.doi.org/10.1016/j.cma.2022.115346, URL https://www.sciencedirect.com/science/article/pii/S0045782522004327. A Special Issue in Honor of the Lifetime Achievements of J. Tinsley Oden.
[51] C. Mora, J.T. Eweis-Labolle, T. Johnson, L. Gadde, R. Bostanabad, Probabilistic neural data fusion for learning from an arbitrary number of multi-fidelity data sets, Comput. Methods Appl. Mech. Engrg. 415 (2023) 116207, http://dx.doi.org/10.1016/j.cma.2023.116207, URL https://www.sciencedirect.com/science/article/pii/S0045782523003316.
[52] K. Kontolati, S. Goswami, M.D. Shields, G.E. Karniadakis, On the influence of over-parameterization in manifold based surrogates and deep neural operators, J. Comput. Phys. 479 (2023) 112008.
[53] X. Meng, H. Babaee, G.E. Karniadakis, Multi-fidelity Bayesian neural networks: Algorithms and applications, J. Comput. Phys. 438 (2021) 110361.
[54] W. Buntine, A. Weigend, Bayesian Back-Propagation, Technical Report FIA-91-22, NASA Ames Research Center Moffett Field, CA, 1991.
[55] H. Ritter, A. Botev, D. Barber, A scalable laplace approximation for neural networks, in: 6th International Conference on Learning Representations, ICLR 2018-Conference Track Proceedings, Vol. 6, International Conference on Representation Learning, 2018.
[56] Z. Deng, F. Zhou, J. Zhu, Accelerated linearized Laplace approximation for Bayesian deep learning, Adv. Neural Inf. Process. Syst. 35 (2022) 2695–2708.

[57] A. Immer, M. Korzepa, M. Bauer, Improving predictions of Bayesian neural nets via local linearization, in: International Conference on Artificial Intelligence and Statistics, PMLR, 2021, pp. 703–711.
[58] A. Immer, M. Bauer, V. Fortuin, G. Rätsch, K.M. Emtiyaz, Scalable marginal likelihood estimation for model selection in deep learning, in: International Conference on Machine Learning, PMLR, 2021, pp. 4563–4573.
[59] M. Humt, Laplace Approximation for Uncertainty Estimation of Deep Neural Networks (Ph.D. thesis), TUM, 2019.
[60] J. Martens, R. Grosse, Optimizing neural networks with kronecker-factored approximate curvature, in: International Conference on Machine Learning, PMLR, 2015, pp. 2408–2417.
[61] A. Botev, H. Ritter, D. Barber, Practical gauss-newton optimisation for deep learning, in: International Conference on Machine Learning, PMLR, 2017, pp. 557–565.
[62] K. Qian, A.S. Liao, S. Gu, V.A. Webster-Wood, Y.J. Zhang, Biomimetic IGA neuron growth modeling with neurite morphometric features and CNN-based prediction, Comput. Methods Appl. Mech. Engrg. 417 (2023) 116213.
[63] L. Pouchard, K.G. Reyes, F.J. Alexander, B.-J. Yoon, A rigorous uncertainty-aware quantification framework is essential for reproducible and replicable machine learning workflows, Digit. Discov. 2 (5) (2023) 1251–1258.
[64] C. Krishnanunni, T. Bui-Thanh, Layerwise sparsifying training and sequential learning strategy for neural architecture adaptation, 2022, arXiv preprint arXiv:2211.06860.
[65] M. Muto, J.L. Beck, Bayesian updating and model class selection for hysteretic structural models using stochastic simulation, J. Vib. Control 14 (1–2) (2008) 7–34.
[66] J.-P. Vila, V. Wagner, P. Neveu, Bayesian nonlinear model selection and neural networks: A conjugate prior approach, IEEE Trans. Neural Netw. 11 (2) (2000) 265–278.
[67] N. Chitsazan, A.A. Nadiri, F.T.-C. Tsai, Prediction and structural uncertainty analyses of artificial neural networks using hierarchical Bayesian model averaging, J. Hydrol. 528 (2015) 52–62.
[68] P. Shekhar, A. Patra, Hierarchical approximations for data reduction and learning at multiple scales, Found. Data Sci. 2 (2) (2020) 123–154.
[69] P. Shekhar, A. Patra, A forward–backward greedy approach for sparse multiscale learning, Comput. Methods Appl. Mech. Engrg. 400 (2022) 115420.
[70] X. Huan, Y.M. Marzouk, Simulation-based optimal Bayesian experimental design for nonlinear systems, J. Comput. Phys. 232 (1) (2013) 288–317.
[71] C. Riis, F. Antunes, F. Hüttel, C. Lima Azevedo, F. Pereira, Bayesian active learning with fully Bayesian Gaussian processes, Adv. Neural Inf. Process. Syst. 35 (2022) 12141–12153.
[72] R. Dehghannasiri, B.-J. Yoon, E.R. Dougherty, Efficient experimental design for uncertainty reduction in gene regulatory networks, in: BMC Bioinform., 16, (13) BioMed Central, 2015, pp. 1–18.
[73] A. Paquette-Rufiange, S. Prudhomme, M. Laforest, Optimal design of validation experiments for the prediction of quantities of interest, 2023, arXiv preprint arXiv:2303.06114.
[74] P.M. Williams, Bayesian regularization and pruning using a Laplace prior, Neural Comput. 7 (1) (1995) 117–143.
[75] S.H. Rudy, T.P. Sapsis, Sparse methods for automatic relevance determination, Physica D 418 (2021) 132843.
[76] E.T. Jaynes, Probability Theory: The Logic of Science, Cambridge University Press, 2003.
[77] E. Daxberger, A. Kristiadi, A. Immer, R. Eschenhagen, M. Bauer, P. Hennig, Laplace redux-effortless Bayesian deep learning, Adv. Neural Inf. Process. Syst. 34 (2021) 20089–20103.
[78] M.S. Alnæs, J. Blechta, J. Hake, A. Johansson, B. Kehlet, A. Logg, C. Richardson, J. Ring, M.E. Rognes, G.N. Wells, The FEniCS project version 1.5, Arch. Numer. Softw. 3 (100) (2015) http://dx.doi.org/10.11588/ans.2015.100.20553.
[79] M. McGurn, D. Salac, ABLATE ablative boundary layers at the exascale, version 0.12.23, 2013, https://ablate.dev.
[80] A. Sarkar, P.K. Singh, L. Zhu, D. Faghihi, S. Ren, Carbon-sequestration straw cellulose-aerogel gradient thermal insulation material, ACS Appl. Eng. Mater. (2024).
[81] L. An, M. Di Luigi, J. Tan, D. Faghihi, S. Ren, Flexible percolation fibrous thermal insulating composite membranes for thermal management, Mater. Adv. 4 (1) (2023) 284–290.
[82] S. Bhattacharjee, Integrating Scientific Machine Learning and Physics-Based Models for Quantification of Uncertainty in Thermal Properties of Silica Aerogel (Ph.D. thesis), State University of New York at Buffalo, 2023.
[83] K.A. Maupin, L.P. Swiler, N.W. Porter, Validation metrics for deterministic and probabilistic data, J. Verif. Valid. Uncertain. Quantif. 3 (3) (2018) 031002.
[84] G. Georgalis, K. Retfalvi, P.E. DesJardin, A. Patra, Combined data and deep learning model uncertainties: An application to the measurement of solid fuel regression rate, Int. J. Uncertain. Quantif. 13 (5) (2023).
[85] G. Surina III, G. Georgalis, S.S. Aphale, A. Patra, P.E. DesJardin, Measurement of hybrid rocket solid fuel regression rate for a slab burner using deep learning, Acta Astronaut. 190 (2022) 160–175.
[86] M. Raissi, P. Perdikaris, G.E. Karniadakis, Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations, J. Comput. Phys. 378 (2019) 686–707.
[87] P.K. Jha, J.T. Oden, Residual-based error corrector operator to enhance accuracy and reliability of neural operator surrogates of nonlinear variational boundary-value problems, Comput. Methods Appl. Mech. Engrg. 419 (2024) 116595, http://dx.doi.org/10.1016/j.cma.2023.116595, URL https://www.sciencedirect.com/science/article/pii/S0045782523007193.
[88] A. Li, Y.J. Zhang, Isogeometric analysis-based physics-informed graph neural network for studying traffic jam in neurons, Comput. Methods Appl. Mech. Engrg. 403 (2023) 115757.
[89] G.E. Box, Robustness in the strategy of scientific model building, Robust. Stat. 1 (1979) 201–236.
[90] G.E. Box, G.C. Tiao, Bayesian Inference in Statistical Analysis, John Wiley & Sons, 2011.
[91] X. Zhou, W. Zhang, H. Xu, T. Zhang, Effective sparsification of neural networks with global sparsity constraint, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 3599–3608.
[92] T. O'Leary-Roseberry, X. Du, A. Chaudhuri, J.R. Martins, K. Willcox, O. Ghattas, Learning high-dimensional parametric maps via reduced basis adaptive residual networks, Comput. Methods Appl. Mech. Engrg. 402 (2022) 115730.
[93] T. O'Leary-Roseberry, U. Villa, P. Chen, O. Ghattas, Derivative-informed projected neural networks for high-dimensional parametric maps governed by PDEs, Comput. Methods Appl. Mech. Engrg. 388 (2022) 114199.
[94] S. Wang, H. Wang, P. Perdikaris, Learning the solution operator of parametric partial differential equations with physics-informed DeepONets, Sci. Adv. 7 (40) (2021) eabi8605.
[95] L. Lu, P. Jin, G.E. Karniadakis, Deeponet: Learning nonlinear operators for identifying differential equations based on the universal approximation theorem of operators, 2019, arXiv preprint arXiv:1910.03193.
[96] Z. Li, N. Kovachki, K. Azizzadenesheli, B. Liu, K. Bhattacharya, A. Stuart, A. Anandkumar, Fourier neural operator for parametric partial differential equations, 2020, arXiv preprint arXiv:2010.08895.
[97] S. Goswami, A. Bora, Y. Yu, G.E. Karniadakis, Physics-informed deep neural operator networks, in: Machine Learning in Modeling and Simulation: Methods and Applications, Springer, 2023, pp. 219–254.