Available online at www.eccomasproceedia.org
Eccomas Proceedia UNCECOMP (2025) 584-596

ECCOMAS
Proceedia

UNCECOMP 2025
6th ECCOMAS Thematic Conference on
Uncertainty Quantification in Computational Science and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou (eds.)
Rhodes Island, Greece, 15-18 June 2025

A HIERARCHICAL BAYESIAN APPROACH FOR POPULATION-BASED SHM IN SHIP HULL STRUCTURES

Georgios Aravanis[1], Nicholas Silionis[2], Jacopo Bardiani[1], Marco Giglio[1], Konstantinos Anyfantis[2], and Claudio Sbarufatti[1]

[1]Politecnico di Milano, Department of Mechanical Engineering
Via La Masa 1, 20156 Milano, Italy
e-mail: {georgios.aravanis, jacopo.bardiani, marco.giglio, claudio.sbarufatti}@polimi.it

[2] National Technical University of Athens, School of Naval Architecture and Marine Engineering
9 Heroon Polytechniou Av., Athens, 15780 Zografos, Greece
e-mail: {nsil,kanyf}@naval.ntua.gr

Abstract. *Structural health monitoring (SHM) strategies involve the processing of structural response data to indirectly assess an asset's condition. These strategies can be enhanced for a group of structures, especially when they are similar, since mutual underlying physics are expected to exist. The concept behind population-based SHM exploits the sharing of data among individuals, so that data-rich members can support data-scarce ones. One approach to population-level modeling is the hierarchical Bayesian method, where the model is structured hierarchically in terms of its parameters, and correlation among learning tasks is enabled by conditioning on shared latent variables.*

This work investigates the application of a hierarchical Bayesian model to infer expected distributions of deflection amplitudes at both the population and domain levels, with the aim of detecting excessive initial deflections in a population of plate elements. Although these damages are typically localized, they can trigger unexpected events, if not properly monitored. The work is conducted in a numerical setting using a Finite Element model to generate strain response data, which serve as the monitoring data. Bayesian inference was conducted using Markov Chain Monte Carlo (MCMC), with a surrogate model employed to calculate the likelihood function. The hierarchical approach was compared to an independent model for a plate component with few data. The results revealed that, under data sparsity conditions, the hierarchical model can offer more robust results in terms of uncertainty, which is essential for decision-making tasks.

Keywords: Hierarchical Bayes, Population-based SHM, Uncertainty quantification Markov chain Monte Carlo, Damage detection.

ISSN:2623-3339 © 2025 The Authors. Published by Eccomas Proceedia.
Peer-review under responsibility of the organizing committee of UNCECOMP 2025.
doi: 10.7712/120225.12379.21105

1 INTRODUCTION

Oftentimes, failure of many engineering assets is associated with localized damage events. In the case of ship hull structures, which are primarily composed of plates and stiffened panels, deterioration of these components can lead to reduced load-carrying capacity and potentially unexpected failure modes. Thus, the deployment of Structural Health Monitoring (SHM) campaigns on-board [1, 2], which are able to monitor such events, becomes highly relevant for helping operators make timely maintenance decisions.

In SHM, the challenge essentially lies in learning Quantities of Interest (QoIs) that characterize the underlying deterioration process. These are most often unobservable, and thus structural response quantities (e.g., strain or acceleration sensor data) are instead used to determine their values. Concurrently, this inference procedure needs to be robust in terms of uncertainty, as the confidence level in the deployed SHM system for decision-making directly depends on how well uncertainty is handled. This problem formulation lends itself well to a Bayesian framework [3], where the goal is to estimate the updated joint probability distribution of the QoIs conditioned on the acquired sensor data. Acknowledging this, numerous studies in the SHM field have employed Bayesian methods to address different tasks in the classical SHM hierarchy [4]; namely, related to (damage) diagnosis [5], identification [6], and prognosis [7, 8].

Despite the benefit of developing SHM approaches for structural integrity management purposes, their effectiveness is bounded by the availability of data relevant to the problem at hand. In real-world large-scale structures, this poses a critical challenge [9], since increased resource restrictions arising from the scale of the system influence data availability. At the same time, their inherent complexity requires sufficiently comprehensive datasets to describe the various phenomena, which one seeks to model. On the other hand, such structures often exhibit fundamental similarities in terms of their constituent components, presenting an opportunity to enable information sharing among them. If we know that they have been produced under similar specification and with the same material, they can be considered nominally identical [10]. Even then however, they are not truly identical due to inherently random events (i.e., relating to manufacturing, assembly, materials, etc.). An example of this is the repeated presence of plate elements in a ship hull. Although produced on the same manufacturing line, they exhibit geometric variability due to fabrication-related factors during the ship's construction, which often manifests as deviations in their intended shape.

The variation in the structural attributes of individual components results in variations in their response, which complicates the development of a model that generalizes well across all of them. Hierarchical Bayesian modeling provides a principled statistical framework to model these variations within a group while leveraging statistical similarities among its members [11, 12], all while providing the key benefit of Bayesian modeling. The statistical correlations enabled through hierarchical models allow for improved inference in sparse data scenarios, whether due to data loss (e.g. transmission issues, sensor failure, etc.) or due to the challenges of online monitoring, where early predictions must be made with limited initial observations. An additional benefit of this approach is that it makes the most out of well-instrumented domains, and thus the overall added value of SHM implementation is increased compared to a more standard approach.

Hierarchical Bayesian models have seen particular success in several works related to population based SHM (PBSHM) [13, 14, 15], as well as other engineering problems framed in a hierarchical manner [16, 17]. This is because of their ability to capture both common trends across a population as well as individual differences among its members, which aligns naturally

with the concept of constructing a general population model. However, despite their benefits these models have yet to be explored in SHM applications for marine structures.

In this work, we demonstrate a hierarchical Bayesian framework on a large-scale double-bottom model of a containership, where damage is define as a set of bathtub-shaped plate deflections The approach involves firs generating realistic strain observations by Finite Element (FE) modeling. These observations are subsequently incorporated into a hierarchical model, alongside an FE-based surrogate, necessary for the calculation of the likelihood function and, ultimately, the prediction of parameters associated with the deflectio amplitudes. Results are presented in terms of variance in the parameter estimates to assess whether data-poor plates can benefi from statistical information at the population level, to enable robust damage detection under data-sparsity.

2 THE DATASET

The dataset used in this work was generated from a detailed FE model, which corresponds to part of the double-bottom structure within the central hold of a containership, and is of size 27 m × 18 m × 2.2 m. A detailed description of the structural information has been omitted here for the sake of brevity but can be found in [18]. This section of the ship hull was selected because it is typically under constant in-plane compressive loading due to the nature of the operational loads a containership experiences. The combination of compression and initial out-of-plane deflection for plate elements can accelerate the onset of buckling, which makes this region particularly critical.

2.1 Finite element modeling particulars

The FE model features 4-node linear shell elements for the external bottom and 2-node linear beam elements to represent the various stiffening members of the double-bottom (i.e., girders, floors longitudinal stiffeners). The stiffness contributed by the inner bottom has been implicitly captured by incorporating an effective width for the girders and floors The group of $K = 6$ plates enclosed within the red boundaries in Figure 1(a) define the region used for strain data generation for the purposes of this study. Each plate, of size $a \times b$, deforms based on a sinusoidal expression given by:

$$w(u, v; \tilde{w}) = \tilde{w} \sin\left(\frac{\pi u}{a}\right) \sin\left(\frac{\pi v}{b}\right) \tag{1}$$

where $\tilde{w}$ denotes the amplitude of the deflection $u \in [0, a]$ and $v \in [0, b]$. Their adjacent plates were also modeled as deflecte to provide realistic states for the surroundings of the monitored region, since all structural elements are expected to exhibit slight imperfections within allowable limits. They are shown with a slight black shading to indicate that they are not part of the data generation process. During the modeling process, $\tilde{w}$ was controlled parametrically to generate different realizations of out-of-plane deflections The perspective shape of a simulated out-of-plane deflectio can be seen in Figure 1(b), where the (conceptual) strain sensor location is also indicated by a red marker. This placement is consistent across all K plates.

The loads considered are those typically experienced by a containership while sailing in calm water conditions; namely, hogging (i.e., negative) hull girder bending, lateral pressure on the side shell and outer bottom, and the cargo weight [19]. At the double-bottom level, these translate into longitudinal (N_x) and transverse (N_y) in-plane compressive stresses, pressure (P) distributed over the hull exterior, and line loads (Q) on the inner bottom, respectively. These were assigned with representative values based on the loading manual of the vessel. In terms

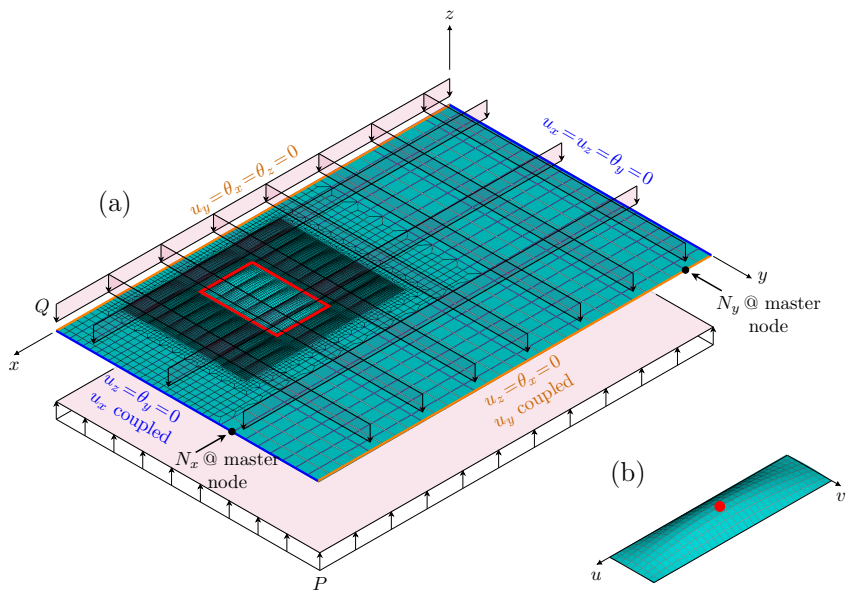

Figure 1: (a) FE model of the subject geometry including the monitored plates (within red boundaries), and (b) representative out-of-plane deflectio with strain sensor at the center. Adapted from the original work of the authors [18].

of the boundary conditions, they were define based on how the considered geometry deforms with respect to the entire structure and are detailed in the descriptions adjacent to the edges of the model.

Regarding the material, linear elastic and isotropic properties (Young's modulus $E = 207$ GPa, Poisson's ratio $v = 0.3$) were assigned. A static, linear elastic analysis was ultimately employed to solve the model, essentially representing conditions when the vessel is in port or sailing in still water (i.e., no waves). Nevertheless, the data sampling rate implied in this context is considered sufficient given the relatively low temporal evolution of out-of-plane deflections

2.2 Dataset generation

In the data generation process, transverse strains ε_{yy} were selected as the observed quantities, which falls in line with the underlying physics of the problem, as plate bending theory dictates that strains in the shorter dimension are the dominant ones. Hereinafter, these are denoted as ε for notational simplicity.

To generate the latent quantities $\tilde{w}$ that give rise to these observations, it is assumed that each plate $k \in K$ is deflecte based on a local (Normal) distribution for $\tilde{w}$, whose parameters are themselves realizations from common higher-level distributions. To this end, global distributions were placed over both the expected value of the deflectio amplitude and its variance. The former captures the variability introduced by fabrication processes, leading to differences in expected deflectio amplitudes among plates. Meanwhile, the latter accounts for changing conditions that introduce additional variability around each expected deflection reflectin uncertainties in the structural response beyond systematic fabrication-induced deviations.

By sampling from these global distributions, K pairs of local expectations and variances were obtained, definin an equal number of local distributions, from each of which N_k samples of $\tilde{w}$ were drawn. The resulting matrix of K-dimensional vectors $\{[\tilde{w}_{i,1}, \ldots, \tilde{w}_{i,k}, \ldots, \tilde{w}_{i,K}]\}_{i=1}^{N_k}$

$\in \mathbb{R}^{N_k \times K}$ was concatenated with additional samples corresponding to the deflection of the neighboring plates, which were assumed to remain within allowable tolerances. The augmented vectors were fed into the constructed FE model to generate transverse strain observations ε, which were subsequently perturbed with additive white Gaussian noise with a standard deviation of 5 $\mu\varepsilon$. The corresponding pairs of $\{\tilde{w}_{i,k}, \varepsilon_{i,k}\}_{i=1}^{N_k}$ for each $k \in K$ are shown with different colors in Figure 2, where $N_k = 20$ for $k \in \{1, \ldots, 5\}$ and $N_k = 2$ for $k = 6$. In this way, some measurements are hidden from the employed models, to demonstrate the robustness of hierarchical modeling in imbalanced dataset settings.

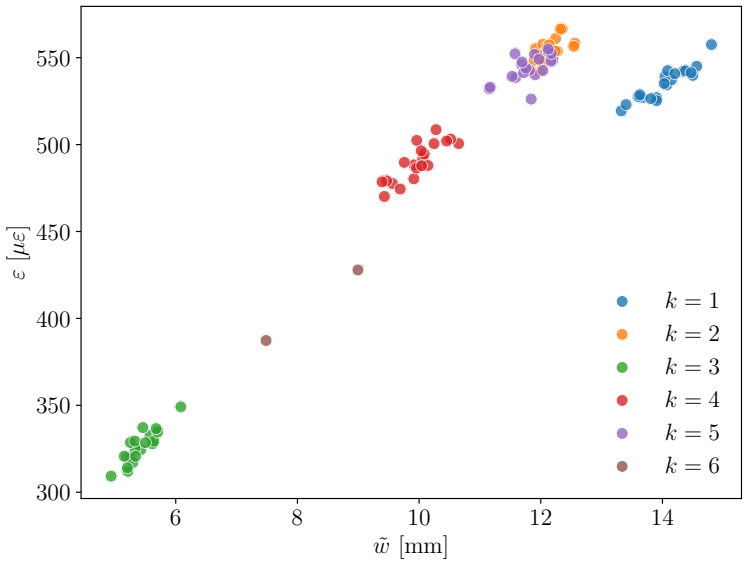

Figure 2: Clusters of generated observations of transverse strain for the six plate components. The brown markers show Plate 6 with scarce data.

3 THE HIERARCHICAL BAYESIAN APPROACH

To establish the hierarchical Bayesian formulation, we represent the data from the entire group of plates as:

$$\{\tilde{\mathbf{w}}_k, \boldsymbol{\varepsilon}_k\}_{k=1}^{K} = \left\{ \{\tilde{w}_{i,k}, \varepsilon_{i,k}\}_{i=1}^{N_k} \right\}_{k=1}^{K} \tag{2}$$

The objective is to learn the posterior distributions of the parameters (local expectations and variances) and hyperparameters (higher-level expectations and variances) that govern the deflectio amplitude $\tilde{w}$, given the observed strain data ε.

3.1 Surrogate modeling

Each plate (domain) is related to its own task, which can be formally expressed as:

$$\varepsilon_{i,k} = \mathcal{M}(\tilde{w}_{i,k}) + \epsilon_{i,k} \tag{3}$$

where $\mathcal{M}(\cdot)$ denotes the model operator and $\epsilon_{i,k}$ is the additive noise term. The former represents the expected mapping from deflection to transverse strain, established by the constructed FE model, which was eventually replaced with a cheaper data-driven surrogate. This is typical practice to avoid computational overhead when employing sampling-based algorithms (i.e.,

Markov Chain Monte Carlo (MCMC)) for the estimation of posterior distributions in a Bayesian context.

The surrogate model was chosen to be a standard Gaussian Process Regression (GPR) model [20] because it guarantees against overfittin and automatically returns confidenc intervals associated with its predictive capacity. The former property is quite relevant in our case for robustly capturing the general pattern of the dataset, as the presence of multiple out-of-plane deflection contributes to some level of variability in the data. The predictive variance provided by the GPR was also analyzed to assess the relative spread of strain values for different deflectio amplitudes (i.e., the coefficien of variation), which was found to be $\sim 7\%$ at the mean deflectio amplitude. Given this level of variance, while acknowledging the computational complexity and inherent high-dimensionality that comes with hierarchical models, it was decided to only take the GPR mean into consideration in the likelihood function evaluation. Figure 3 illustrates a trained GPR model for a specifi plate $k \in K$. Independent surrogate models were trained for each task to prevent bias in the mean prediction, and thus the model operator $\mathcal{M}(\cdot)$ in Equation 3 is more pertinently denoted as $\mathcal{M}_k(\cdot)$ for clarity.

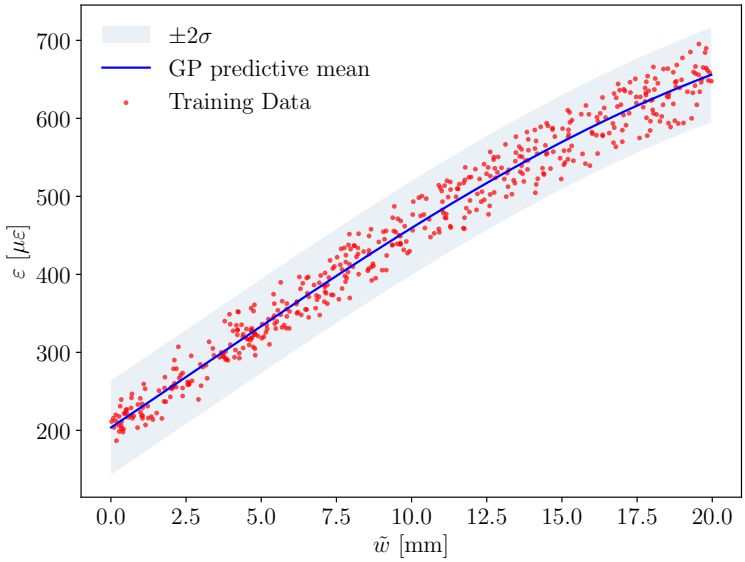

Figure 3: GPR surrogate model for Plate 6. Red markers represent the training-set (FE test data), the blue line shows the posterior predictive mean, and the shaded region indicates a 2σ credible area. Note that for the construction of the likelihood function we use only the mean of the GPR.

3.2 The hierarchical model

This work compares a partially-pooled hierarchical model, which assumes correlation between plate-specifi parameters, with an independent (no-pooling) model that does not, to motivate sharing information within the group of plates. Pooling refers to the way information is shared across members in hierarchical models; a no-pooling approach predicts separate parameters for each member, while conversely, a complete pooling approach assumes all members share the same parameters. Partial pooling strikes a balance by allowing some parameters to vary across members and correlation is enabled among them by conditioning on shared latent variables [3]. Since the deflectio amplitudes related to each plate are assumed to be sampled from distributions with correlated parameters for the partially-pooled hierarchical model, one

expects that the data-rich members ($k = 1, \ldots, 5$) should support the data-scarce one ($k = 6$).

The hierarchical Bayesian model structure employed in this work is now presented. This follows from the work of Brealy et al. [15]. The likelihood function for the model was constructed using the commonly employed zero-mean Gaussian measurement error model,

$$\left\{ \{\varepsilon_{i,k}\}_{i=1}^{N_k} \right\}_{k=1}^{K} \sim \mathcal{N}(\mathcal{M}_k(\tilde{w}_{i,k}), \gamma^2) \tag{4}$$

where γ^2 is the variance of the measurement error term ϵ. This was considered to be the same for each plate under the assumption that the same hardware were used for data acquisition.

To start the Bayesian formulation, one can set prior distributions over the deflectio amplitude realizations $\tilde{w}_{i,k}$ related to each plate:

$$\left\{ \{\tilde{w}_{i,k}\}_{i=1}^{N_k} \right\}_{k=1}^{K} \sim \mathcal{N}(\mu_{\tilde{w}_k}, \sigma^2_{\tilde{w}_k}) \tag{5}$$

The expectation $\mu_{\tilde{w}_k}$ and variance $\sigma^2_{\tilde{w}_k}$ associated with each plate are themselves sampled from global Normal distributions,

$$\{\mu_{\tilde{w}_k}\}_{k=1}^{K} \sim \mathcal{N}(\mu_\mu, \sigma^2_\mu) \tag{6}$$

$$\{\sigma_{\tilde{w}_k}\}_{k=1}^{K} \sim \mathcal{N}(\mu_\sigma, \sigma^2_\sigma) \tag{7}$$

where μ_μ, σ^2_μ, μ_σ and σ^2_σ are the higher-level parameters, which are shared among plates. One should also encode prior knowledge of the learning tasks as prior distributions over these higher-level parameters, referred to as hyperpriors. All four were assigned Gamma hyperpriors, more specifically

$$\mu_\mu \sim \mathcal{G}(\text{shape} = 3, \text{rate} = 0.2) \tag{8}$$

$$\sigma_\mu \sim \mathcal{G}(\text{shape} = 0.8, \text{rate} = 0.35) \tag{9}$$

$$\mu_\sigma \sim \mathcal{G}(\text{shape} = 3.6, \text{rate} = 6) \tag{10}$$

$$\sigma_\sigma \sim \mathcal{G}(\text{shape} = 4.8, \text{rate} = 16) \tag{11}$$

The Gamma distribution was chosen to ensure positivity, which is essential for the physical parameters involved in the model. The constants in Equations (8)-(11) (given in mm) were specifie using the formulas for the two basic statistical moments (i.e., expected value and variance) of a Gamma-distributed random variable θ. These are given by [3],

$$\mathrm{E}[\theta] = \frac{\text{shape}}{\text{rate}} \tag{12}$$

$$\mathrm{Var}[\theta] = \frac{\text{shape}}{\text{rate}^2} \tag{13}$$

Following this, the shape and rate parameters were chosen in a way that the resulting variance for each higher-level parameter allows for sufficien exploration across a reasonable range of values. With the same rationale, the expected value was deliberately set to deviate to a certain extent from the ground-truthvalues used in data generation, ensuring that the model actually learns from the data. We also ensured that the chosen hyperprior parameters led to parameter/hyperparameter samples that, ultimately, determined non-negative deflectio amplitudes, despite using Normal distributions.

Similarly to the higher-level parameters, a Gamma prior distribution was set for the shared noise variance γ^2,

$$\gamma \sim \mathcal{G}(\text{shape} = 80, \text{rate} = 16) \tag{14}$$

where the chosen parameter values reflec a fairly narrow support of possible values, as practically in most cases measurement noise is known a priori. Note that hypepriors for γ are given in $\mu\varepsilon$.

Figure 4(a) shows a directed acyclic graph (DAG) of the Bayesian model under a hierarchical approach. The plate-level parameters are indexed by k, and plate notation is used to indicate that these nodes are repeated over the K plate components. Shaded and unshaded nodes represent observed and latent variables, respectively. Higher-level nodes are positioned outside the plates and are not indexed by k, denoting that they are shared across all members. Figure 4(b) illustrates the DAG for the independent approach. The independent models can be derived by removing the K plate from the hierarchical model's DAG and explicitly indexing the noise variance by k. Although the plate-level parameters are still conditioned on shared higher-level nodes, in this case, these nodes are only informed by the data from one specifi member.

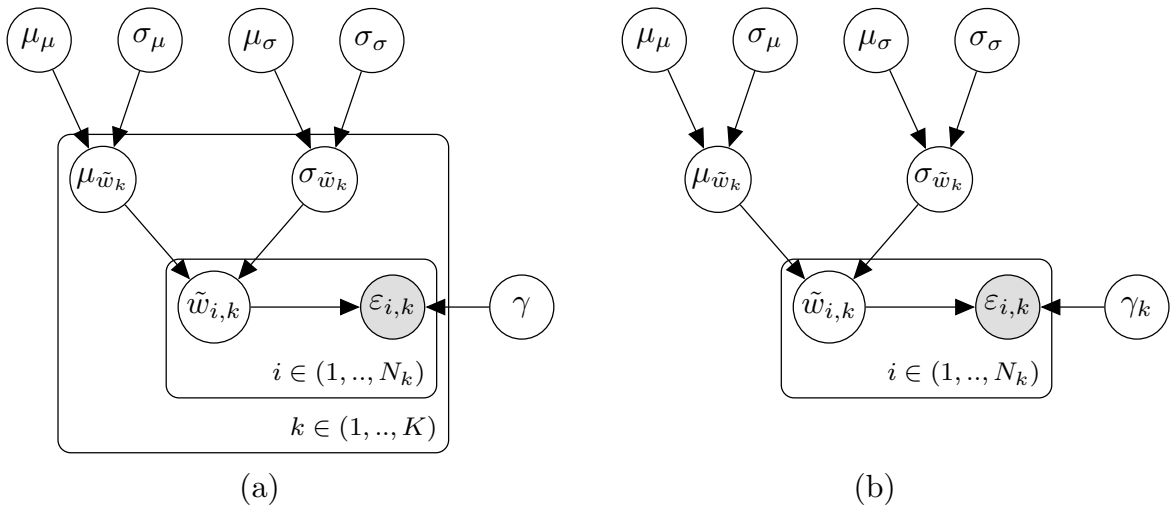

Figure 4: DAGs representing (a) the hierarchical (partial-pooling) model and (b) the independent model.

4 RESULTS

4.1 Inference

Inference of the parameters was performed using MCMC methods, via the no U-turn variant (NUTS) of Hamiltonian Monte Carlo (HMC) [21]. The implementation was carried out

in `NumPyro` [22], which uses automatic differentiation to make gradient-based samplers like NUTS feasible.

The setup comprised of 4000 samples for warmup and 2000 samples for inference, run across four parallel chains, with each chain starting from a randomly generated initial state to ensure convergence. The rank-normalized $\hat{R}$ diagnostic [23] was employed as quantitative check of convergence, and posteriors were accepted when this was lower than 1.01. This was satisfie for all parameters/hyperparameters, as can be seen by the trace plots on the right-hand side of Figure 5. The resulting kernel density function (KDE)-based posteriors for all chains (shown by each line style) are also shown on the left-hand side of the figure The hyperpriors for the shared parameters have been plotted with magenta dashed lines, while the color scheme for the plate-specifi parameters follows the convention used in Figure 2.

From the results, one notices that the modes of the posterior distributions for the shared higher-level parameters (apart from γ) have shifted away from their prior modes. Additionally, the variance of these posterior distributions is notably reduced. These observations provide assurance that the algorithm is working properly and that the model has learned from the measurements. As expected, the posterior distribution of the noise variance γ remains relatively unchanged from its prior due to the tightly constrained prior range and the inherent properties of the Gamma distribution, which features shallow tails that might affect the NUTS sampler during exploration of the posterior. However, this is compensated via μ_σ which encodes any additional variance observed in the data, and behaves as intended.

For the plate-specifi parameters $\mu_{\tilde{w}_k}$ and $\sigma_{\tilde{w}_k}$, the posterior variance for the data-scarce plate (i.e., $k = 6$) is considerably larger than that of the data-rich plates. This is expected, given that it is informed by only two observations, whereas the other plates have 20 data points each. A comparison was conducted between this posterior distribution and the one obtained when modeling this plate independently (i.e., without pooling information from the broader dataset) in order to show the benefi of hierarchical models.

4.2 Damage detection

While the inferred latent variables related to deflectio levels are those that describe the underlying structural behavior, they are not directly observed in an operational setting. Instead, for downstream tasks, such as damage detection, one should revert to the strain space, as this is what can be monitored and used by operators.

Along these lines, the posterior distributions of the plate-specifi parameters were used for damage detection in the following way. First, posterior MCMC samples were taken for the expectation and variance for plate $k = 6$, to draw a deflectio amplitude each time from a Normal distribution with the same statistics as the parameter draws (according to Equation 5). Next, to obtain posterior predictive samples for the transverse strain, the drawn deflectio amplitude samples were evaluated through the mean function of $\mathcal{M}_{k=6}(\cdot)$, and the likelihood function was applied using the posterior predictive measurement noise γ^2 (see Equation 4), for each sample. KDE estimation was then employed to approximate the posterior predictive distribution of strains, using the samples from the likelihood. This process was applied for both the hierarchical (partial-pooling) and the independent (no-pooling) approaches.

The posterior predictive distributions for the transverse strain are shown in Figure 6, with the vertical lines indicating strain values associated with different deflectio thresholds set by maritime regulatory bodies. Namely, the 8 mm level corresponds to a limit deflectio threshold that can be seen as a damage indicator, while the 4 mm represents a standard level expected under normal operating conditions. The 6 mm level constitutes an intermediate state. The

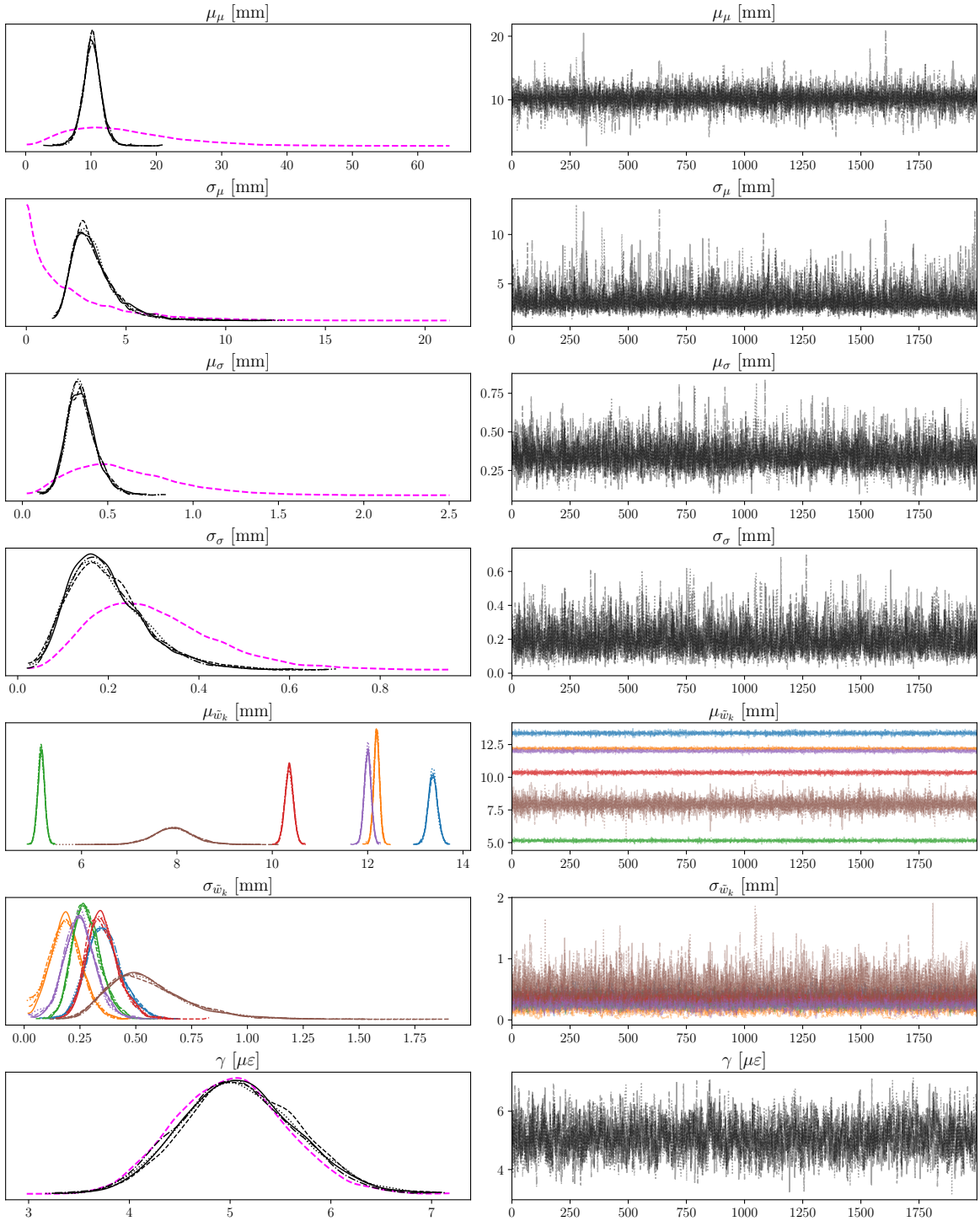

Figure 5: KDE-based posteriors of hierarchical model parameters/hyperparameters (left panel) and trace plots (right panel). Dashed magenta lines show the prior distributions applied to the shared parameters. The color scheme for $k = \{1, \ldots, 6\}$ follows: blue, orange, green, red, purple, and brown.

strain values for these thresholds were computed by passing them into the surrogate model an equal number of times as the posterior MCMC samples of γ^2, drawing observations as per the likelihood, and then averaging the results for each deflectio level.

It is evident that the partial pooling model has managed a noticeable reduction in the level of

uncertainty in its predictions compared to the no-pooling model. This follows from the fact that the parameters pertaining to the latent variable have been learned within a joint inference setting. Such a reduction in uncertainty is particularly valuable in industrial applications, where high uncertainty can lead to operational challenges. For instance, if a deflectio amplitude exceeding the 8 mm threshold implies inspection, large uncertainties could result in higher probability of false alarms to occur. This, in turn, would lead to unnecessary downtime and increased costs for stakeholders. At the same time, within a signal detection theory context [24], the reduction in uncertainty enhances the probability of detection by improving the model's ability to distinguish actual damage from normal variations. This leads to a more reliable assessment of structural integrity and reduces the risk of missed detections.

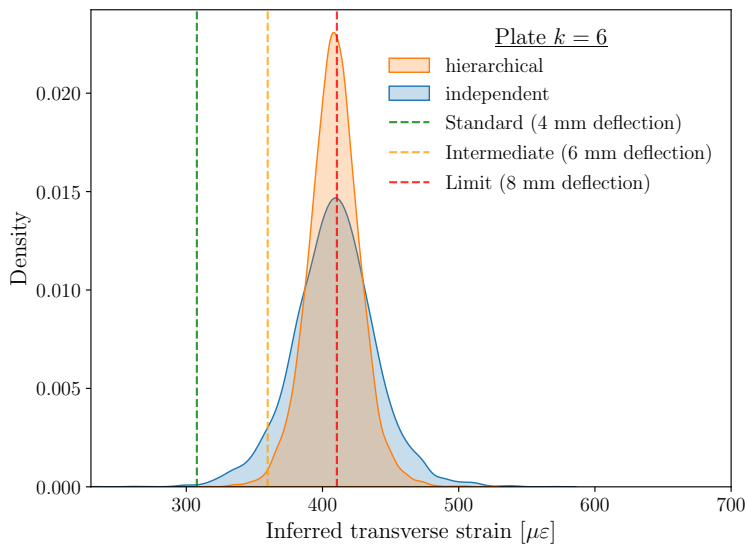

Figure 6: KDE-based posterior predictive densities of strain under a hierarchical (partial-pooling) approach and an independent (no-pooling) approach. The dashed lines represent averaged strain values for different deflectio levels.

5 CONCLUSIONS

In this work, a hierarchical Bayesian model has been constructed to learn the latent (both local and global) parameters that drive a fabrication-induced deterioration phenomenon commonly encountered in ship hull structures, i.e., out-of-plane deflections A set of deflecte plates was assumed in a region of the double-bottom structure of a containership, and synthetic strain observations were generated through an FE model. These observations served as features in the inference process, where an MCMC sampler, specificall the NUTS algorithm, was employed. Given the computational demands of NUTS, a surrogate model was used in place of the FE model to improve computational efficien y.

To demonstrate the capabilities of hierarchical modeling, a specifi plate was assumed with less data compared to the others. The hierarchical model enabled this data-poor plate to borrow statistical strength from the broader population, reducing prediction uncertainty. The refine posterior uncertainties of the parameters were compared against an independent no-pooling approach, with results showing that the hierarchical model enables more informed decision-making in data sparse scenarios.

REFERENCES

[1] C. R. Farrar and K. Worden, *Structural Health Monitoring: A Machine Learning Perspective.* John Wiley & Sons, 2012.

[2] C. R. Farrar and K. Worden, "An introduction to structural health monitoring", *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 365, no. 1851, pp. 303–315, 2007.

[3] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin. *Bayesian Data Analysis.* CRC press, third edition, 2013.

[4] A. Rytter, "Vibrational based inspection of civil engineering structures", Ph.D.dissertation, Aalborg University, Denmark, 1993.

[5] P. Gardner, L.A. Bull, N. Dervilis, K. Worden, "On the application of kernelised Bayesian transfer learning to population-based structural health monitoring", *Mechanical Systems and Signal Processing*, vol. 167, p. 108519, 2022.

[6] M. K. Ramancha, M. A. Vega, J. P. Conte, M. D. Todd, and Z. Hu, "Bayesian model updating with finit element vs surrogate models: Application to a miter gate structural system," *Engineering Structures*, vol. 272, p. 114901, 2022.

[7] D. Cristiani, C. Sbarufatti, F. Cadini, and M. Giglio, "Fatigue damage diagnosis and prognosis of an aeronautical structure based on surrogate modelling and particle filte ," *Structural Health Monitoring*, vol. 20, no. 5, pp. 2726–2746, 2021.

[8] D. Cristiani, C. Sbarufatti, and M. Giglio, "Damage diagnosis and prognosis in composite double cantilever beam coupons by particle filterin and surrogate modelling," *Structural Health Monitoring*, vol. 20, no. 3, pp. 1030-1050, 2021.

[9] P. Cawley, "Structural health monitoring: Closing the gap between research and industrial deployment", *Structural Health Monitoring*, vol. 17, no. 5, pp. 1225-1244, 2018.

[10] L.A. Bull, P.A. Gardner, J. Gosliga, T.J. Rogers, N. Dervilis, E.J. Cross, E. Papatheou, A.E. Maguire, C. Campos, and K. Worden, "Foundations of population-based SHM, Part I: Homogeneous populations and forms", *Mechanical Systems and Signal Processing*, vol. 148, p. 107141, 2021.

[11] A. Gelman and J. Hill. *Data Analysis Using Regression and Multilevel/ Hierarchical Models*. Cambridge university press, 2006.

[12] I. G. Kreft and J. De Leeuw. *Introducing Multilevel Modeling*. Sage, 1998.

[13] L. A. Bull, D. Di Francesco, M. Dhada, O. Steinert, T. Lindgren, A. K. Parlikad, A. B. Duncan and M. Girolami, "Hierarchical Bayesian modeling for knowledge transfer across engineering fleet via multitask learning", *Computer-Aided Civil and Infrastructure Engineering*, vol. 38, no. 7, pp. 821–848, 2023.

[14] T.A. Dardeno, K. Worden, N. Dervilis, R.S. Mills and L.A. Bull, "On the hierarchical Bayesian modelling of frequency response functions", *Mechanical Systems and Signal Processing*, vol. 208, p. 111072, 2024.

[15] S. M. Brealy, A. J. Hughes, T. A. Dardeno, L. A. Bull, R. S. Mills, N. Dervilis, and K. Worden, "Multitask learning for improved scour detection: A dynamic wave tank study", *arXiv preprint arXiv:2408.16527*, 2024.

[16] N. Papadimas and T. Dodwell, "A hierarchical Bayesian approach for calibration of stochastic material models," *Data-Centric Engineering*, vol. 2, p. e20, 2021.

[17] D. Di Francesco, M. Chryssanthopoulos, M. H. Faber, and U. Bharadwaj, "Decision-theoretic inspection planning using imperfect and incomplete data," *Data-Centric Engineering*, vol. 2, p. e18, 2021.

[18] G.I. Aravanis, N.E. Silionis, K.N. Anyfantis, "Damage detection in ship hull structures under operational variability through strain sensing", *Ocean Engineering*, vol. 286, p. 115537, 2023.

[19] Y. Sumi *et al.*, *Final report of committee on large container ship safety*. Tokyo, Japan: Committee on large container ship safety, Maritime Bureau of Japan's Ministry of Land, Infrastructure, Transport and Tourism (MILT), 2015.

[20] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)*. The MIT Press, 2005

[21] M. D. Hoffman, and A. Gelman, "The No-U-Turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo", *Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1593–1623, 2014.

[22] D. Phan, N pradhan and M. Jankowiak, "Composable effects for fl xible and accelerated probabilistic programming in NumPyro", *arXiv preprint arXiv:1912.11554*, 2019.

[23] A. Vehtari, A. Gelman, D. Simpson, B. Carpenter, and P.-C. Bürkner, "Rank-Normalization, Folding, and Localization: An Improved $\hat{R}$ for Assessing Convergence of MCMC (with Discussion)," *Bayesian Analysis*, vol. 16, no. 2, 2021.

[24] S.M. Kay. *Fundamentals of Statistical Signal Processing Volume II: Detection Theory*. Prentice Hall, 1998.