Article:
Möller, S., Jones, M.R., Jonscher, C. et al. (2026) Grey-box Gaussian processes for mode shape normalisation: damage localisation under environmental and operational variability. Mechanical Systems and Signal Processing, 253. 114353. ISSN: 0888-3270

https://doi.org/10.1016/j.ymssp.2026.114353

Mechanical Systems and Signal Processing 253 (2026) 114353

Contents lists available at ScienceDirect

Mechanical Systems and Signal Processing

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

Full length article

Grey-box Gaussian processes for mode shape normalisation: Damage localisation under environmental and operational variability

Sören Möller [a,*], Matthew R. Jones [b], Clemens Jonscher [a], Jasper Ragnitz [a], Elizabeth J. Cross [b], Raimund Rolfes [a]

[a] Leibniz Universität Hannover, Institute of Structural Analysis, Appelstr. 9A, 30167 Hannover, Germany
[b] Dynamics Research Group, School of Mechanical, Aerospace and Civil Engineering, University of Sheffield, Mappin Street, Sheffield S1 3JD, UK

ARTICLE INFO

Communicated by E. Chatzi

Keywords:
Model-based SHM
EOVs
Mode shapes
Gaussian process regression
Grey-box modelling

ABSTRACT

Structural health monitoring (SHM) offers a promising path towards automated, long-term monitoring of critical infrastructure such as offshore structures and bridges. A crucial component of such monitoring schemes is the localisation of damage, with model-based SHM providing a possible framework for this task. Here, damage localisation can be achieved by minimising the discrepancy between the modal properties of a damaged state and those of a healthy reference. However, this process is severely hindered by changes in the modal properties induced by environmental and operational variations (EOVs). This difficulty is compounded by the fact that measurement data are often limited: data may be unavailable across the full operational span, may be missing because of sensor failure and dropout, or may be only sparsely sampled because of hardware constraints. In particular, mode shapes are often used without adequately accounting for the limited coverage of EOVs in continuous model-based damage localisation frameworks, resulting in inaccurate localisation.

This paper proposes a regression-based data normalisation scheme that learns mode shapes as functions of spatial coordinate and EOVs using Gaussian process regression, allowing them to vary across a structure's operating envelope and thereby enabling model-based damage localisation under varying environmental and operational conditions. To further alleviate the problem of limited training data, two grey-box modelling strategies based on Gaussian processes that incorporate accessible engineering knowledge are considered: (i) a Hilbert-space Gaussian process enforcing boundary conditions, and (ii) a Gaussian process with a physics-based prior mean given by finite-element mode shapes. The focus is on long-term monitoring scenarios in which training data are limited and purely data-driven regression techniques perform poorly. Using the Leibniz University Test Structure for Monitoring, a representative real-world structure, this study demonstrates that incorporating accessible engineering-domain knowledge into Gaussian process models alleviates data scarcity and improves damage localisation under partially observed EOVs. These results highlight the practical value of the proposed grey-box models for SHM. They enable mode-shape normalisation for model-based damage localisation under environmental and operational variability, even when training data are limited, thereby enhancing continuous, long-term monitoring of operational structures.

* Corresponding author.
E-mail address: s.moeller@isd.uni-hannover.de (S. Möller).

https://doi.org/10.1016/j.ymssp.2026.114353
Received 19 November 2025; Received in revised form 26 March 2026; Accepted 27 April 2026
Available online 30 April 2026

1. Introduction

Long-term, vibration-based structural health monitoring (SHM) identifies damage by continuously analysing a structure's measured dynamic responses [1]. In many monitoring scenarios, it is desirable to go beyond damage detection and also locate its position. In particular, for large-scale structures, providing maintenance engineers with the location of damage can save considerable time and reduce costs. Among the available vibration-based techniques, vibration-based model updating, as an inverse method, offers a principled route to damage localisation [2]. The central premise is that changes in a structure's geometry or material properties manifest themselves as changes in its dynamic responses [3]. Consequently, the experimentally identified modal properties, namely the scalar natural frequencies and the vector-valued mode shapes, no longer align with those predicted by the finite-element (FE) model of the undamaged structure [4]. The differences between measurement data and the FE model can be quantified using an objective function and reduced by treating them as an optimisation problem [5]. Here, the model's stiffness parameters are iteratively adjusted to restore agreement with the measured observations, and the updated model provides quantitative insight into the probable location and severity of each stiffness reduction. Günther et al. [6] introduced a relative least-squares error metric, which is also employed in the present study. This metric eliminates the mismatch between measurement and the numerical model by considering changes relative to their respective damaged and healthy states. For example applications of this metric in model-based SHM, see Wolniak et al. [7] (single-objective) and Ragnitz et al. [8] (multi-objective).

Unfortunately, environmental and operational variations (EOVs) can affect the structure's behaviour and mask damage-related changes in the modal properties [9]. Many authors have investigated the influence of EOVs on the natural frequencies of different structures, including bridges [10,11], bell towers [12], and wind turbine towers [13]. While previous studies demonstrate that EOVs heavily influence natural frequencies, mode shapes are generally considered to be more robust. Early studies by Li et al. [14] and Ni et al. [15] on cable-stayed bridges reported no significant influence of EOVs such as temperature and wind speed on the identified mode shapes, and Xia et al. [16] similarly observed no clear correlation between mode shapes and temperature or humidity for an outdoor reinforced-concrete slab, attributing this to EOVs acting approximately uniformly on the structure. However, Ni et al. [15] also noted that variations in the mode shapes due to EOVs might be masked by measurement errors, making it difficult to quantify the true impact of EOVs. More recent work provides stronger evidence that EOVs can systematically affect mode shapes: Xia et al. [17] showed that decreasing temperatures reduce the *modal assurance criterion* (MAC) values at the Guangzhou New TV Tower, indicating a correlation between temperature and mode shapes, attributable to non-uniform temperature distributions. Additionally, a study by Jonscher et al. [18] revealed a correlation between mode shapes and operational parameters, specifically the nacelle orientation of a 3.4 MW onshore wind turbine. Similar observations have also been reported for smaller-scale structures: Wernitz et al. [19] reported temperature-dependent variations in the MAC for the Leibniz University Test Structure for Monitoring (LUMO), a 9 metre high lattice tower. Collectively, these studies demonstrate that EOVs can cause non-negligible, systematic variability in mode shapes. This poses a particular challenge for model-based SHM approaches, which, as noted by Farrar et al. [20], still struggle to account for the EOVs inherent to almost all in-situ structures. For example, they showed that a nominally symmetric, north–south-oriented bridge exhibits markedly different first bending modes in the morning and early evening due to different sun heating of its opposing sides [10,20]. Such behaviour illustrates how EOV-induced variability complicates the selection of suitable reference mode shapes for model updating.

To avoid erroneous damage localisation, it is therefore essential to normalise the modal properties with respect to EOV-induced effects, a process commonly referred to as *data normalisation*. Early studies mitigated EOV-induced effects for model updating by either aligning measurements to a reference state [21–23] or incorporating temperature dependence directly into the FE model [24–26], with the latter, however, requiring significant computational resources, which limits its applicability to real-time-capable SHM [27]. Because the objective function in this study employs a relative metric, data normalisation is required only for the measured modal properties. The FE model itself does not need to be adjusted to account for EOVs; only calibration to the healthy state is necessary. To embed data normalisation within a relative-metric objective function, a regression-based approach is required, in which the structural response is modelled explicitly as a function of the relevant EOVs [28–30]. A principal challenge is extending this approach to vector-valued mode shapes. In the context of model updating, Ragnitz et al. [23] proposed a clustering approach that groups mode shapes with similar EOV regimes, thereby enhancing the accuracy of the damage localisation. However, to the authors' knowledge, no prior work has mitigated EOV-induced variations in mode shapes using a regression-based approach suitable for continuous model updating in the context of long-term, real-time-capable SHM. Such a regression-based approach significantly reduces the need for manual intervention (e.g., the choice of the number of clusters and cluster assignment) during model updating. Moreover, it enables the use of purely data-driven (black-box) models (e.g., vanilla neural networks) to extract complex relationships between EOVs and mode shapes directly from the data.

Whilst purely data-driven techniques are attractive for learning mode shapes, they are known to extrapolate poorly when EOV regimes are not represented in the training set. In the case of complete physical understanding, a physics-based (white-box) model can be applied. Because the model's governing equations are fully explicit, its internal mechanics are transparent, enabling it to make predictions in EOV regimes that are not included in the training data. At the same time, fully characterising EOV-induced effects with white-box approaches is often challenging; for example, formulating first-principles models of the nonlinear behaviour at very low temperatures (e.g., due to variations in soil stiffness and material properties) in closed form is difficult. It is, therefore, reasonable to combine data-driven modelling (black box), which can extract hidden relationships from the measurement data with knowledge of the underlying physics (white box), to create a grey-box model capable of producing reliable predictions for previously unseen EOVs. A general overview of different kinds of grey-box models can be found in [31–33].

In the context of SHM, Cross et al. [34] demonstrated through several case studies that Gaussian process (GP) regression is a suitable tool for grey-box modelling. As Bayesian nonparametric models, GPs offer several desirable properties for this setting. For the data-driven part, this includes the ability to capture a wide range of behaviours without being limited to a specific parametric model, requiring only a few *a priori* inputs, and handling noisy and uncertain observations [35]. For the physics-based part, prior knowledge can be embedded via the prior mean [36,37], kernel design [38,39], or general physical constraints (e.g., boundary conditions, monotonicity, convexity, inequality constraints) [40–43].

Regarding data normalisation, recent work has shown that grey-box models can successfully compensate for EOVs in natural frequencies [44,45]. By contrast, grey-box data normalisation of mode shapes for model-based damage localisation under environmental and operational variability remains less established. This study addresses this gap by considering grey-box strategies for regression-based data normalisation that model the reference mode shape in the relative least-squares objective explicitly as a function of sensor position and EOVs, thereby aiming to improve robustness under limited training data and EOV coverage. From an engineering standpoint, prior knowledge of boundary conditions and an FE model are often available for the structure under investigation, offering two opportunities to embed domain knowledge. This study presents two grey-box models to support mode shape prediction under sparse or incomplete training data and limited EOV coverage:

(i) **Physics-based boundary condition**: The displacement boundary condition for the clamped end of a mode shape $\boldsymbol{\varphi}$ can be enforced by adopting a reduced-rank Hilbert-space GP whose eigenfunctions satisfy $\boldsymbol{\varphi}(x = 0) = 0$ by construction, where x denotes a spatial coordinate.

(ii) **Physics-based prior mean**: The mode shape of an FE model of the structure is used as the GP's prior mean, allowing the data-driven component to model only the residual variability. Consequently, EOV-induced effects, which are challenging to model explicitly in the FE model, are learned from the data, whereas the FE mode shape supplies solely spatial information.

To evaluate the performance of the proposed methods, several analysis scenarios are investigated that reflect key challenges in the practical deployment of SHM on operational structures. In this work, data are obtained from the Leibniz University Test Structure for Monitoring (LUMO) – an outdoor lattice used to benchmark SHM methods [19]. The first analysis scenario considers the situation in which training data are available across the relevant EOV range, but only as a limited number of sparse measurements. In smartphone-based SHM applications [46], for example, it is desirable to minimise the number of measurement intervals required for training data collection while still achieving adequate coverage of the full EOV range. The second addresses limited measurement locations, due to budget constraints or sensor failures. The third evaluates the models' performance on previously unseen EOVs, as rare weather events are not represented in the training data. Finally, a prescribed damage scenario at LUMO is considered to assess continuous, model-based damage localisation using mode shapes alone, thereby isolating the impact of mode shape data normalisation. When employing the relative least-squares error metric, or any other discrepancy metric, reliable localisation is linked to accurate mode shape predictions across the range of EOVs, which, in turn, is crucial for reasonable maintenance decision-making.

A core focus of the work presented here is to identify the type of model that offers the best performance for predicting mode shapes of operational structures, thereby offering practical guidance for SHM practitioners. The analysis considers data-driven approaches as well as grey-box models that utilise available engineering knowledge for the task at hand. Discussion is offered on how the different forms of physical knowledge can assist in mode shape prediction, including scenarios where the available physics is not directly representative of behaviour outside the training data.

The remainder of the study is organised as follows. Section 2 introduces a regression-based data normalisation scheme to capture EOV effects on mode shapes and sets a black-box baseline via GP regression. Two strategies are then proposed for embedding engineering knowledge into GPs. Section 3 presents LUMO and quantifies its temperature-dependent mode shape variations. Section 4 examines the grey-box approaches under sparse and incomplete spatial coverage of the mode shapes. Furthermore, the robustness of each model to previously unseen EOVs is evaluated, and the role of physical knowledge in mode-shape prediction is discussed. Section 5 shows how a grey-box model can be integrated into a continuous, model-based damage-localisation framework on LUMO using mode shapes alone, with the training data covering only a limited range of EOVs.

2. Grey-box models

Gaussian process (GP) regression serves as the black-box model used in this study. It models the relationship between inputs $\boldsymbol{x}$ and targets $\boldsymbol{y}$ through a latent function $f(\boldsymbol{x})$, accounting for additive Gaussian noise, such that $y = f(\boldsymbol{x}) + \epsilon$, with $\epsilon \sim \mathcal{N}(0, \sigma_n^2)$. Given training inputs $\boldsymbol{X} = [\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n]$ and targets $\boldsymbol{y}$, predictions at new points $\boldsymbol{x}_*$ are obtained using Bayesian inference. A GP is solely defined by its mean function $m(\boldsymbol{x})$, typically set to zero, and its covariance (kernel) function $k(\boldsymbol{x}, \boldsymbol{x}')$:

$$f(\boldsymbol{x}) \sim \mathcal{GP}(m(\boldsymbol{x}), k(\boldsymbol{x}, \boldsymbol{x}')). \tag{1}$$

The kernel function quantifies input similarity and determines smoothness. A common choice is the squared-exponential (SE) kernel:

$$k_{\mathrm{SE}}(\boldsymbol{x}, \boldsymbol{x}') = \sigma_f^2 \exp\left(-\frac{\|\boldsymbol{x} - \boldsymbol{x}'\|^2}{2\ell^2}\right), \tag{2}$$

where σ_f^2 is the signal variance and ℓ the length scale. The SE kernel is employed throughout the present study. For a comprehensive treatment of GP theory, see Rasmussen et al. [35].

Prior knowledge can be incorporated into a GP in several ways, yielding a spectrum of grey-box representations. The more physics-based information that is embedded (provided it is representative of the problem), the less the model must rely on data alone. Hence, a model lies towards the white end of the spectrum when extensive prior knowledge is supplied and towards the black end when only minimal prior knowledge is used [34]. For example, a GP with a zero-mean prior and an SE kernel is a purely data-driven black-box model, written as $\mathcal{GP}(0, k_{\mathrm{DATA}})$, where the subscript DATA indicates the data-driven nature of the kernel. If the mean behaviour can be stated analytically, for example, via governing differential equations, a physics-based prior $m_{\mathrm{P}}(\mathbf{x})$ can be supplied, providing a light-grey model $\mathcal{GP}(m_{\mathrm{P}}, k_{\mathrm{DATA}})$. Here, m_{P} captures the known physics P, while the kernel represents the residual variability [34]. When explicit governing equations cannot be derived, one can instead embed general physical constraints (e.g., boundary, monotonicity, convexity, or inequality conditions) directly into the GP. Such constraints require far less detailed physics than differential equations and are therefore easier to formulate and implement. Incorporating constraints positions the model towards the dark-grey end of the spectrum: physics-informed yet still reliant on data, with the SE kernel modelling any remaining unexplained variability. There are numerous ways to apply constraints [40]. The present study focuses on constrained GPs for learning mode shapes of a lattice tower. The constraints considered are physical boundary conditions, specifically the zero-displacement boundary condition of the form $\boldsymbol{\varphi}(0) = 0$ when predicting a mode shape $\boldsymbol{\varphi}$. In this context, Hilbert-space GPs (HSGP) [47] are examined, in which the covariance is reformulated as a Laplacian eigenfunction expansion, implicitly satisfying the boundary condition $\boldsymbol{\varphi}(0) = 0$. In this study, the HSGP is used exclusively with Dirichlet boundary conditions, following Solin et al. [47]. Although mixed boundary conditions (e.g., Dirichlet/Neumann or Robin) are, in principle, admissible, they lie outside the scope of this work and are left for future research. As an alternative way to incorporate the boundary condition $\boldsymbol{\varphi}(0) = 0$, this study also employs a straightforward virtual-point approach, denoted by $\mathcal{GP}_{\mathrm{VP}}(0, k_{\mathrm{DATA}})$, which enforces the zero-displacement boundary condition by augmenting the training dataset $\mathcal{D}$ with noiseless pseudo-measurements equal to zero at the clamped end, yielding the augmented training dataset $\tilde{\mathcal{D}}$. However, this approach might not always be the best option since it can only approximate surface constraints using a finite set of virtual points, and the number of virtual points grows exponentially with input dimension – an instance of the *curse of dimensionality*.

2.1. Hilbert-space GP

By defining eigenfunctions through boundary-value problems, HSGPs naturally incorporate physically meaningful constraints, eliminating the need for artificially imposed boundary observations [47]. HSGPs utilise reduced-rank approximations to efficiently represent covariance structures and facilitate the incorporation of constraints [48]. The central concept is the approximation of the full Gram matrix $\boldsymbol{K}$ by a reduced-rank matrix $\tilde{\boldsymbol{K}}$, where the rank is $m < n$, m being the number of basis functions and n the number of data points. The optimal reduced-rank approximation of $\boldsymbol{K}$, with respect to the Frobenius norm, is $\tilde{\boldsymbol{K}} = \boldsymbol{\Phi}\boldsymbol{\Lambda}\boldsymbol{\Phi}^{\top}$, where $\boldsymbol{\Lambda}$ contains the leading m eigenvalues of $\boldsymbol{K}$, and $\boldsymbol{\Phi}$ consists of the corresponding orthonormal eigenvectors.

For stationary covariance functions, where the covariance $k(\boldsymbol{x}, \boldsymbol{x}')$ depends only on the separation $\boldsymbol{r} = \boldsymbol{x} - \boldsymbol{x}'$, the Wiener–Khintchin theorem provides an integral representation through the spectral density $S(\omega)$:

$$\begin{aligned} k(\boldsymbol{r}) &= \frac{1}{(2\pi)^d} \int S(\omega) e^{\mathrm{i}\omega^{\top}\boldsymbol{r}}\, d\omega, \\ S(\omega) &= \int k(\boldsymbol{r}) e^{-\mathrm{i}\omega^{\top}\boldsymbol{r}}\, d\boldsymbol{r}. \end{aligned} \tag{3}$$

Specifically, for the SE kernel with length scale ℓ and magnitude σ_f^2, the spectral density is given by:

$$S_{\mathrm{SE}}(\omega) = \sigma_f^2 (2\pi\ell^2)^{d/2} \exp\left(-\frac{1}{2}\ell^2 \|\omega\|^2\right). \tag{4}$$

The eigenfunctions $\phi_j(\boldsymbol{x})$ and eigenvalues λ_j used for the kernel approximation come from solving the Laplace eigenvalue problem with Dirichlet boundary conditions:

$$\begin{cases} -\nabla^2 \phi_j(\boldsymbol{x}) = \lambda_j \phi_j(\boldsymbol{x}), & \boldsymbol{x} \in \Omega, \\ \phi_j(\boldsymbol{x}) = 0, & \boldsymbol{x} \in \partial\Omega. \end{cases} \tag{5}$$

These eigenfunctions form an orthonormal basis, enabling the kernel approximation:

$$k(\boldsymbol{x}, \boldsymbol{x}') \approx \sum_{j=1}^{m} S(\sqrt{\lambda_j})\, \phi_j(\boldsymbol{x})\, \phi_j(\boldsymbol{x}') = \boldsymbol{\Phi}\boldsymbol{\Lambda}\boldsymbol{\Phi}^{\top}, \tag{6}$$

with:

$$\boldsymbol{\Phi}_i = \left(\phi_1(\boldsymbol{x}_i), \phi_2(\boldsymbol{x}_i), \ldots, \phi_m(\boldsymbol{x}_i)\right), \tag{7}$$

$$\boldsymbol{\Lambda} = \mathrm{diag}\left(S(\sqrt{\lambda_1}), S(\sqrt{\lambda_2}), \ldots, S(\sqrt{\lambda_m})\right), \tag{8}$$

for $i = 1, \ldots, n$. For Gaussian likelihoods, predictions at a new input $\boldsymbol{x}_*$ are efficiently approximated by:

$$\begin{aligned} \mathbb{E}[f_*] &\approx \boldsymbol{\Phi}_*(\boldsymbol{\Phi}^{\top}\boldsymbol{\Phi} + \sigma_n^2 \boldsymbol{\Lambda}^{-1})^{-1} \boldsymbol{\Phi}^{\top} \boldsymbol{y}, \\ \mathbb{V}[f_*] &\approx \sigma_n^2 \boldsymbol{\Phi}_*(\boldsymbol{\Phi}^{\top}\boldsymbol{\Phi} + \sigma_n^2 \boldsymbol{\Lambda}^{-1})^{-1} \boldsymbol{\Phi}_*^{\top}. \end{aligned} \tag{9}$$

This reduced-rank approach significantly improves computational efficiency, scaling as $\mathcal{O}(nm^2)$ for predictions and $\mathcal{O}(m^3)$ for hyperparameter learning, compared to the standard $\mathcal{O}(n^3)$ complexity of traditional GP methods. Regarding the choice of m, Solin et al. [47] showed that the HSGP converges to the full GP solution when the number of leading eigenfunctions, m, is large enough to resolve the short-length-scale behaviour. For the work conducted in this paper, 1024 eigenfunctions are used [47].

2.2. Physics-based prior mean

If parts of the mean behaviour can be expressed analytically, this knowledge can be encoded directly in the GP's prior mean. Such a model is written as $\mathcal{GP}(m_\mathrm{P}, k_\mathrm{DATA})$, where the physics-based prior m_P is complemented by the SE kernel to capture additional variability. A key advantage is that m_P need not conform to any specific functional form, making this approach widely applicable across different domains [49,50]. In regions of sparse data coverage, the grey-box model $\mathcal{GP}(m_\mathrm{P}, k_\mathrm{DATA})$ relies more heavily on m_P, thereby improving predictive accuracy. One way to incorporate physical knowledge into the prior mean is residual modelling, which replaces the original targets $\boldsymbol{y}$ with the residual $\boldsymbol{\delta} = \boldsymbol{y} - m_\mathrm{P}(\boldsymbol{X})$ and the GP is denoted by $\mathcal{GP}_\delta(0, k_\mathrm{DATA})$. The model predictions are given by:

$$\boldsymbol{y} = m_\mathrm{P}(\boldsymbol{X}) + \mathcal{GP}_\delta(0, k_\mathrm{DATA}). \tag{10}$$

In this study, the physics-based prior mean m_P is given by the analytical mode shape obtained from an FE model of the structure. This analytical mode shape is independent of EOVs and encodes only spatial information. The EOV-induced effects, which are difficult to model explicitly in the FE model, are learned by the data-driven component of $\mathcal{GP}(m_\mathrm{P}, k_\mathrm{DATA})$.

2.3. Implementation and model updating procedure

The grey-box models[1] considered in this study are used for regression-based data normalisation of mode shapes to enable model-based damage localisation under EOVs and limited training data. Their implementation is as follows. For each mode shape, a single GP is trained on all available mode-shape amplitude observations across the sensor network. For each training sample, the inputs are the sensor position and the relevant EOVs, and the output is the modal amplitude at that position under those EOV conditions. Fig. 1 illustrates this setup.

With this formulation, the modal amplitudes at different sensor positions are modelled within a single GP through the sensor-position component of the input. Hence, the dependence of the modal amplitude on sensor position is captured through the covariance kernel defined over the inputs.

During training, the inputs are scaled to the interval $[-0.5, 0.5]$. For all models except the HSGP, the targets are mean-centred to enable a zero-mean GP formulation. By contrast, the HSGP is trained on the original targets because of the imposed boundary condition $\boldsymbol{\varphi}(0) = 0$. Kernel hyperparameters (cf. Eqs. (2) and (4)) are learned by minimising the negative log marginal likelihood [35,48] using the L-BFGS optimiser.

As a subsequent step towards model-based damage localisation, model updating is performed in the deterministic setting of Ragnitz et al. [8]. Here, deterministic refers only to the optimisation stage of the model updating, where the *global pattern search algorithm* [52] is employed. As a deterministic optimisation method, it yields reproducible results for identical inputs. The objective is defined as a relative error in the coordinate-wise mode shape components and is closely related to the *enhanced coordinate modal assurance criterion* [53], thereby leveraging the full information content of the mode shapes. The relative metric between the undamaged (labelled 0) and damaged (labelled 1) states of the simulated (S) and measured (M) data is defined as

$$\epsilon_{\boldsymbol{\varphi}} = \sum_{k=1}^{N_\mathrm{modes}} \left\| \left(\boldsymbol{\varphi}_{S_1,k}(\boldsymbol{\Xi}) - \boldsymbol{\varphi}_{S_0,k} \right) - \left(\boldsymbol{\varphi}_{M_1,k} - \boldsymbol{\varphi}_{M_0,k} \right) \right\|_2^2, \tag{11}$$

where N_modes is the number of modes considered and $\boldsymbol{\Xi}$ is a design parameter set [6]. Eq. (11) comprises two parts: the first represents the relative change in mode shapes predicted by the model being updated, while the second is based on the mode shapes extracted from measurements using BayOMA. To incorporate the grey-box predictions into the objective, the reference mode shapes $\boldsymbol{\varphi}_{M_0,k}$, $k = 1, \ldots, N_\mathrm{modes}$ are replaced by the corresponding grey-box predictions $\hat{\boldsymbol{\varphi}}_{M_0,k}$.

3. Application case study: LUMO

The Leibniz University Test Structure for Monitoring (LUMO), depicted in Fig. 2(a), is a 9 m high steel lattice tower comprising three identical 3 m sections, weighing approximately 90 kg [19]. A key advantage of this test structure is that reversible local damage can be imposed: at six heights (DAM1–DAM6), up to three struts (weighing 55 g each) can be loosened or removed to induce local stiffness and mass changes (Fig. 2(b)). Eighteen uniaxial IEPE accelerometers are mounted in orthogonal pairs at nine measurement levels (ML1–ML9) to capture spatial motion (Fig. 2(b)).

Data are sampled at 1651.61 Hz and stored in 10-min blocks. Although temperature may vary along the LUMO's height, the material temperature is measured only at the clamped end using a Pt100 sensor. Throughout this study, temperature refers to this measurement. Analysis is performed for three different conditions: Stiffening due to freezing (StE), healthy (H), and damaged (DAM6). These three states are listed in Table 1.

[1] All models, except the HSGP, are implemented with GPJax [51]

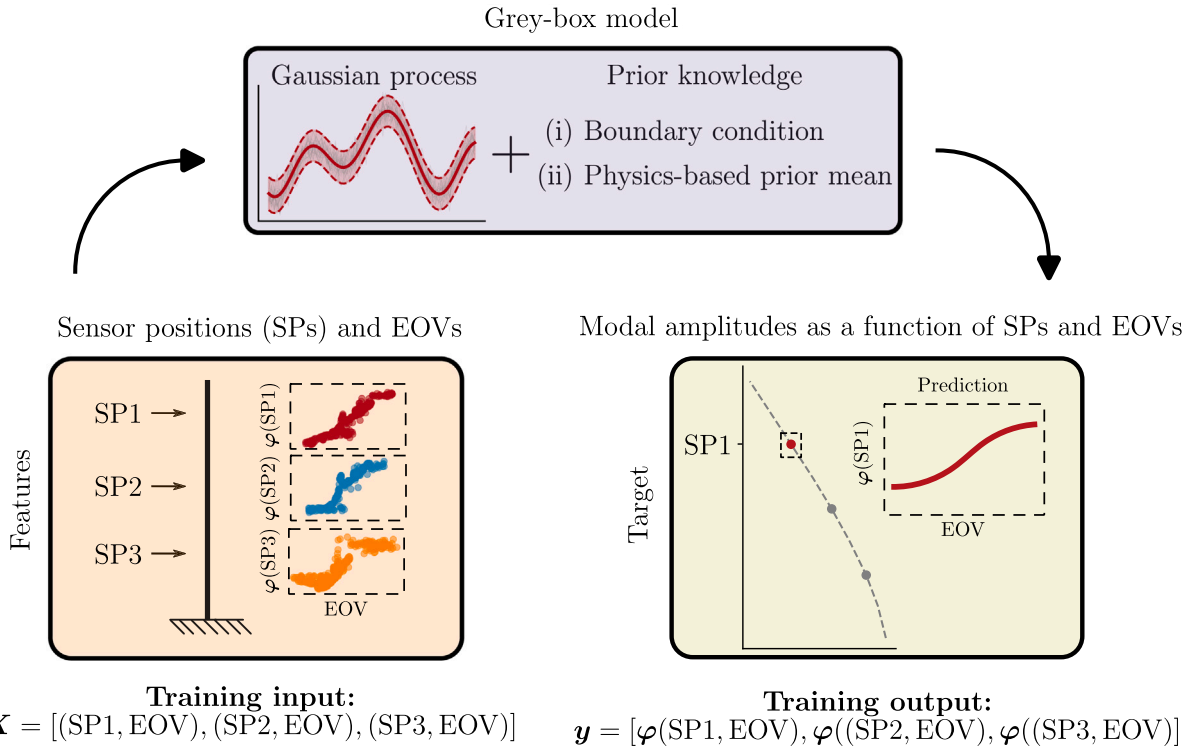


Fig. 1. Prediction scheme for mode shape $\boldsymbol{\varphi}$. The grey-box model takes sensor positions (SPs) and EOVs as inputs and predicts the EOV-dependent modal amplitude $\boldsymbol{\varphi}(\mathrm{SP}, \mathrm{EOV})$ at each sensor location. A grey-box model is trained jointly on data from all sensor positions, with the corresponding EOVs as additional inputs.

Table 1
States of the structure and corresponding labels.

Condition of the structure	Period of data acquisition	Label
Stiffening effect	01.02.2021 to 08.02.2021	StE
Healthy	20.04.2021 to 04.05.2021	H
Damaged - one strut removed	04.05.2021 to 18.05.2021	DAM6

Table 2
Comparison of MAC values for the first and second bending modes (B1-y, B1-x, B2-y, B2-x) for the states StE, H, and DAM6 (cf. Table 1). $\boldsymbol{\varphi}$ denotes the vector-valued mode shape.

Metric	B1-y	B1-x	B2-y	B2-x
$\mathrm{MAC}(\boldsymbol{\varphi}_{\mathrm{H}}, \boldsymbol{\varphi}_{\mathrm{StE}})$	0.9996	0.9998	0.9995	0.9998
$\mathrm{MAC}(\boldsymbol{\varphi}_{\mathrm{H}}, \boldsymbol{\varphi}_{\mathrm{DAM6}})$	1.0000	1.0000	0.9025	1.0000

For state StE, stiffening associated with freezing is assumed to induce nonlinear interactions between temperature and mode shapes [54]. State DAM6 corresponds to the damage location at position DAM6, shown in Fig. 2(b). Here, one strut is entirely removed (cf. Fig. 2(c)).

Following Ragnitz et al. [8,23], this study focuses on the mode shapes of the first two bending mode pairs. The mode shapes are extracted from measurement data using *Bayesian operational modal analysis* (BayOMA); a detailed description of the identification procedure is provided by Jonscher et al. [55]. The identified mode shapes were aligned to arbitrarily chosen reference shapes via a least-squares rotation, correcting for deviations in their vibration directions (cf. [8,23,55]). Fig. 3 shows example mode shapes identified with BayOMA from the measurement data for the states listed in Table 1.

It can be observed that mode shape B2-y (Fig. 3(c)) responds markedly to the removal of a strut (DAM6). For state StE, the mode shapes in Fig. 3 appear unchanged. Nevertheless, the *modal assurance criterion* (MAC) values in Table 2 reveal that the mode shapes B1-y, B1-x, and B2-x of state StE deviate more from state H than those of state DAM6 differ from state H.

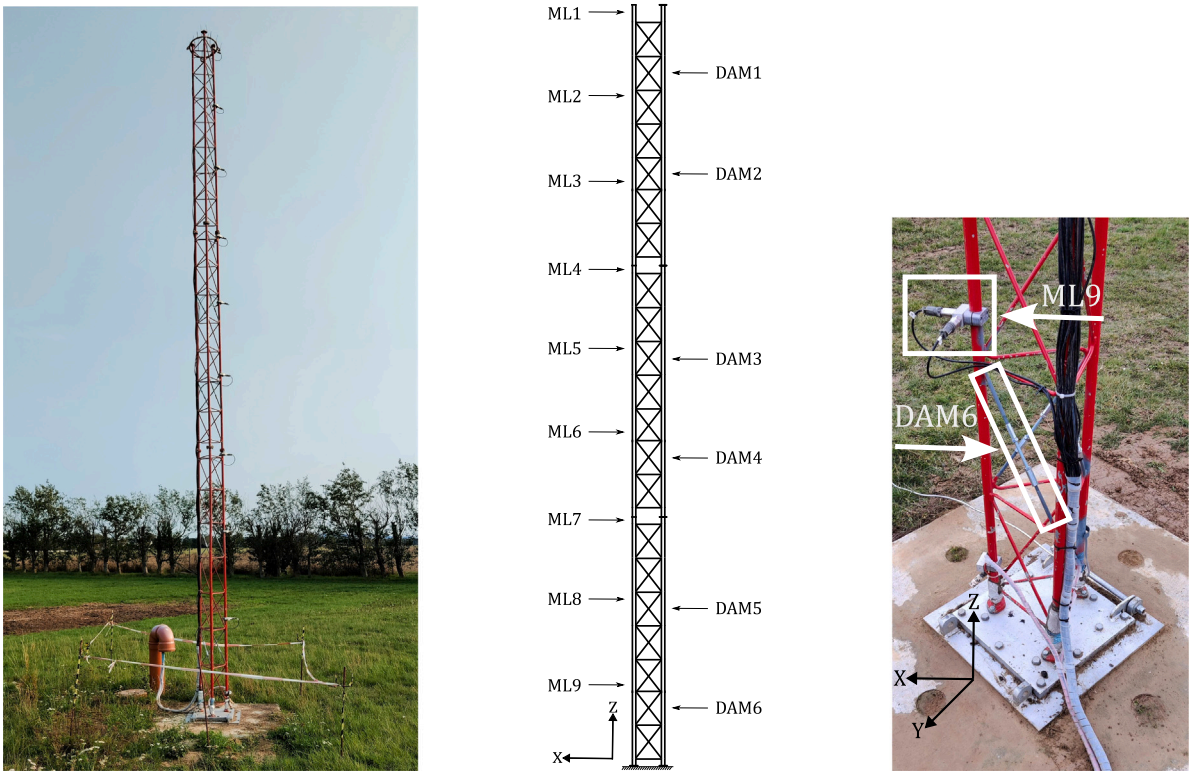


Fig. 2. Photograph of the LUMO structure (2(a)), a schematic of the structure highlighting the nine acceleration measurement levels (ML) and the damage locations (DAM) (2(b)), and the damage location 6 (DAM6) displaying the reversible damage mechanisms as well as the sensors at measurement level 9 (ML9) in the x and y directions (2(c)).

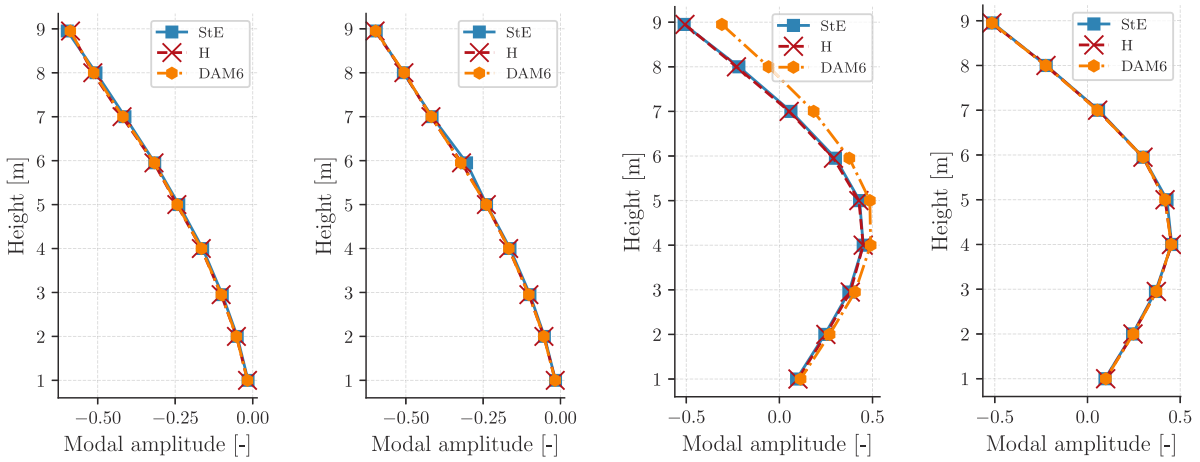
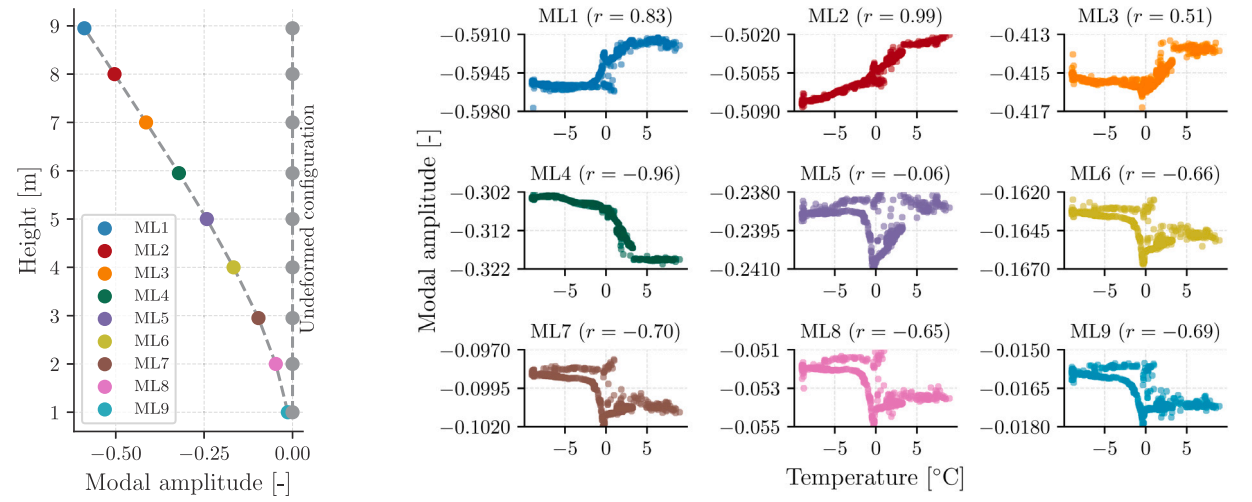


Fig. 3. The first two bending modes in the x- (3(b) and 3(d)) and y-directions (3(a) and 3(c)) for all three states considered. Depicted are the first identified mode shapes for state H and DAM6, while for state StE, the mode shape recorded at −9 °C is presented.

An analysis of the individual mode amplitudes at each ML in Fig. 4 shows that the observed discrepancy is driven by temperature dependence, including a step change at approximately 0°C. This change is not negligible. For instance, at ML1, the modal amplitude changes by an average of 0.604% when the temperature moves from positive to negative values. In contrast, it changes by only 0.097% between the healthy and damaged conditions in state DAM6 (one entire strut removed).



(a) Mode shape B1-x (b) Temperature dependency of modal amplitudes

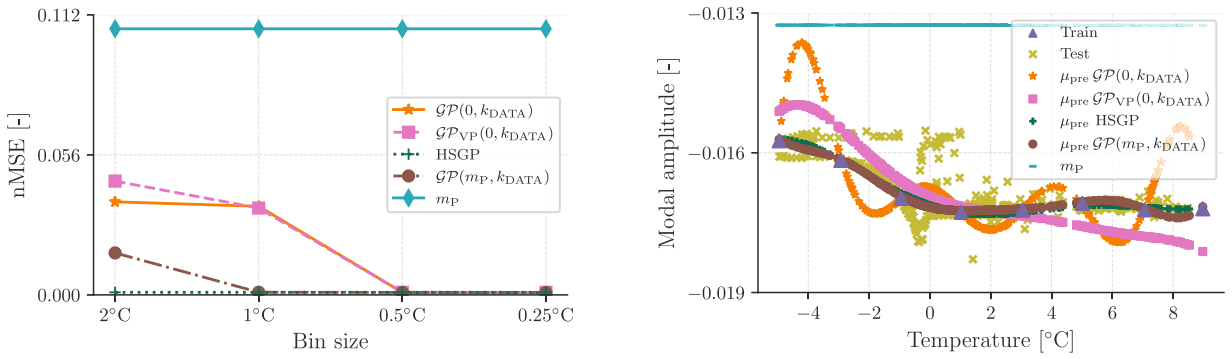
Fig. 4. Temperature-dependent variations of the BayOMA-identified modal amplitudes for bending mode B1-x. (4(a)) estimated mode shape and undeformed configuration. (4(b)) Scatter plots of modal amplitudes versus temperature for each ML, including Spearman's rank correlation coefficient r.

In this study, Spearman's rank correlation coefficient r is used to quantify the monotonic relationship between modal amplitude and temperature; values of +1 or −1 indicate a perfect monotonic increase or decrease, respectively. Using B1-x as an example, a clear relationship is observed across all MLs, despite the behaviour at ML5 (Fig. 4(b)). Two distinct trajectories for the modal amplitude are evident at ML3 and ML5-ML9, with the effect being most pronounced at ML5. The bifurcation is attributed to changes in soil stiffness: freezing and subsequent thawing affect the soil differently. This results in different structural responses and the formation of separate branches. Accordingly, the separation between the two trajectories is most pronounced at the lower measurement levels (ML5–ML9). It should be emphasised that the bifurcation behaviour is present only in the BayOMA-identified modal amplitudes used for training and is not imposed as prior knowledge or explicitly encoded in the considered grey-box formulations. In Fig. 4(b), it can also be observed that the modal amplitudes at different MLs vary in opposite directions, for example, ML1 versus ML4. This behaviour stems from the unit-norm constraint imposed on each mode shape identified using BayOMA. If one amplitude increases, another must decrease so that the overall shape retains unit length. In summary, temperature, particularly the transition region around 0°C, strongly affects the investigated mode shapes, in line with the findings of Wernitz et al. [19]. It is therefore treated as the primary environmental driver in the subsequent regression-based data normalisation on LUMO. However, selecting relevant environmental variables is application-specific and will differ across other structures and SHM scenarios.

In this study, a single grey-box model is used to predict the mode shape $\boldsymbol{\varphi}$ across all measurement levels (cf. Section 2.3). The model takes ML height and temperature as inputs and predicts modal amplitudes as a function of both variables. Consequently, modal amplitudes at different MLs are not treated as independent outputs; rather, the dependence of modal amplitude on ML height is captured through the covariance kernel over the input variables. The physics-based prior mean is obtained from an FE model implemented in DeSiO, an in-house simulation tool [56]. The structural subsystem follows a multi-body finite-element framework composed of rigid bodies and geometrically exact beams. Geometric and material properties of LUMO are adopted from Wernitz et al. [19]; consistent with that work, the brace-leg welds and the flange connections between the segments are modelled as inflexible joints. The resulting FE model comprises 216 geometrically exact beams, with the base fixed rigidly to the ground. To validate the model, the first two bending mode shapes were extracted from measurement data using BayOMA and compared with the corresponding FE mode shapes using the MAC, yielding values of 0.999, 0.999, 0.996, and 0.995 for the first two bending mode pairs, indicating a high level of agreement.

4. Learning mode shapes with limited training data

Using data from LUMO, the proposed grey-box models introduced in Section 2 are evaluated through three representative analysis scenarios for practical SHM. The first scenario considers situations in which only limited training data are available. This setting is common in smartphone-based SHM applications, where measurements are recorded only by a smartphone [46]. In such cases, sparse training data arise when an operator wishes to gain insight into the health of a structure shortly after deployment, while still requiring good coverage of the EOV range, such that the full range of conditions must be sampled efficiently. A similar situation occurs with wireless or battery-powered sensors that can acquire only a few readings per day. Alternatively, an operator may be interested in monitoring a system over a limited but rare subset of environmental conditions, such as the transition between negative and positive temperatures. The second analysis scenario considers cases with limited measurement locations on the monitored



(a) nMSE for different bin sizes (b) Prediction at ML9 for a bin size of 2°C

Fig. 5. Results for the prediction of mode shape B1-x for sparse temperature measurements: (5(a)) shows the nMSE for different bin sizes and (5(b)) depicts the predicted modal amplitudes corresponding to ML9 for a temperature bin size of 2 °C.

structure. This may arise when a monitoring team has a limited budget and can deploy only a small number of sensors, or when sensors have failed and are yet to be replaced, but predictions are still required at those locations. The final scenario, arguably the most challenging, addresses the case where EOVs fall outside the range observed during training. This may occur when predictions are required in operational or environmental regimes not covered by the training data, such as during rare weather events, or when the objective is long- or infinite-horizon health monitoring. In all three analysis scenarios, data from State StE (Table 1) are used to represent a practically relevant SHM setting in which the temperature-related transition around 0 °C strongly affects the modal amplitudes, only a limited number of trials is available and temperatures well below 0 °C are poorly represented.

For the remainder of this study, "predictions" refer to the GP posterior mean μ_{pre} unless stated otherwise.

4.1. Sparse temperature measurements

The first analysis scenario evaluates the methods' ability to reconstruct LUMO mode shapes from sparsely sampled training data at different temperatures, as the representative EOV, while still covering the full temperature range. This mirrors SHM practice, where an operator may be interested only in a narrow environmental regime, for example, the transition across 0 °C. Accordingly, in this analysis scenario, the temperature range in the training data is restricted to $[-5\ °C, 9\ °C]$, capturing the 0 °C transition that strongly affects modal amplitudes whilst keeping the training window narrow. To introduce sparsity into the training set, the measurements from all measurement levels are grouped into temperature bins of progressively finer width (2 °C, 1 °C, 0.5 °C, and 0.25 °C). For a quantitative assessment, Fig. 5(a) presents the normalised mean-squared error (nMSE) of mode shape B1-x for each temperature-bin width. The metric is defined as:

$$\mathrm{nMSE} = \frac{100}{N\sigma_y^2}(\boldsymbol{y}_* - \boldsymbol{y})^{\top}(\boldsymbol{y}_* - \boldsymbol{y}), \tag{12}$$

where N is the number of points, σ_y^2 is the variance of the true targets $\boldsymbol{y}$, and $\boldsymbol{y}_*$ are the model predictions. Here, a score of 0 indicates perfect agreement between the model predictions and the targets, whereas a score of 100 indicates that the model predicts the mean.

Fig. 5(a) reveals that using only the FE-mode-shape prior mean m_{P} is the worst-performing approach. The incorporation of measurement data is therefore beneficial in all cases. For a temperature-bin width of 2 °C, the HSGP achieves the best predictive accuracy, which can be attributed to the explicit enforcement of the physical boundary condition, which introduces limited but reliable prior knowledge. By contrast, under highly sparse training-data coverage (bin size > 1°C), the data-driven component of $\mathcal{GP}(m_{\mathrm{P}}, k_{\mathrm{DATA}})$ provides only limited corrective power, such that the posterior remains biased towards the imperfect prior mean m_{P}. Consequently, $\mathcal{GP}(m_{\mathrm{P}}, k_{\mathrm{DATA}})$ attains a comparably low error, but remains slightly worse than the HSGP at a bin size of 2 °C. At the same bin width, $\mathcal{GP}_{\mathrm{VP}}(0, k_{\mathrm{DATA}})$ performs slightly worse than the black-box baseline $\mathcal{GP}(0, k_{\mathrm{DATA}})$, an effect attributed to the added virtual training points, which inflate the effective training set and exacerbate the curse of dimensionality. With sufficiently dense training-data coverage (bin size $\lesssim 0.5$°C), all models achieve comparable predictive accuracy. To illustrate these trends at the level of individual sensor locations, Fig. 5(b) shows the predicted modal amplitudes at ML9, the measurement level closest to the clamped end, for a temperature-bin width of 2 °C, corresponding to an extremely sparse training set. Since the available prior knowledge is purely spatial and does not encode temperature effects, neither grey-box formulation reproduces the two distinct trajectories that emerge during the step change. Moreover, the associated change in the characteristic length-scale cannot be captured by a standard SE kernel. Local approaches, such as change-point kernels [57] or mixture-of-experts models [58], could offer improvements; however, they require substantially more observational data to ensure a good model fit. As the purpose of this study is to investigate potential solutions for dealing with training data restrictions, these complex models are not pursued further here.

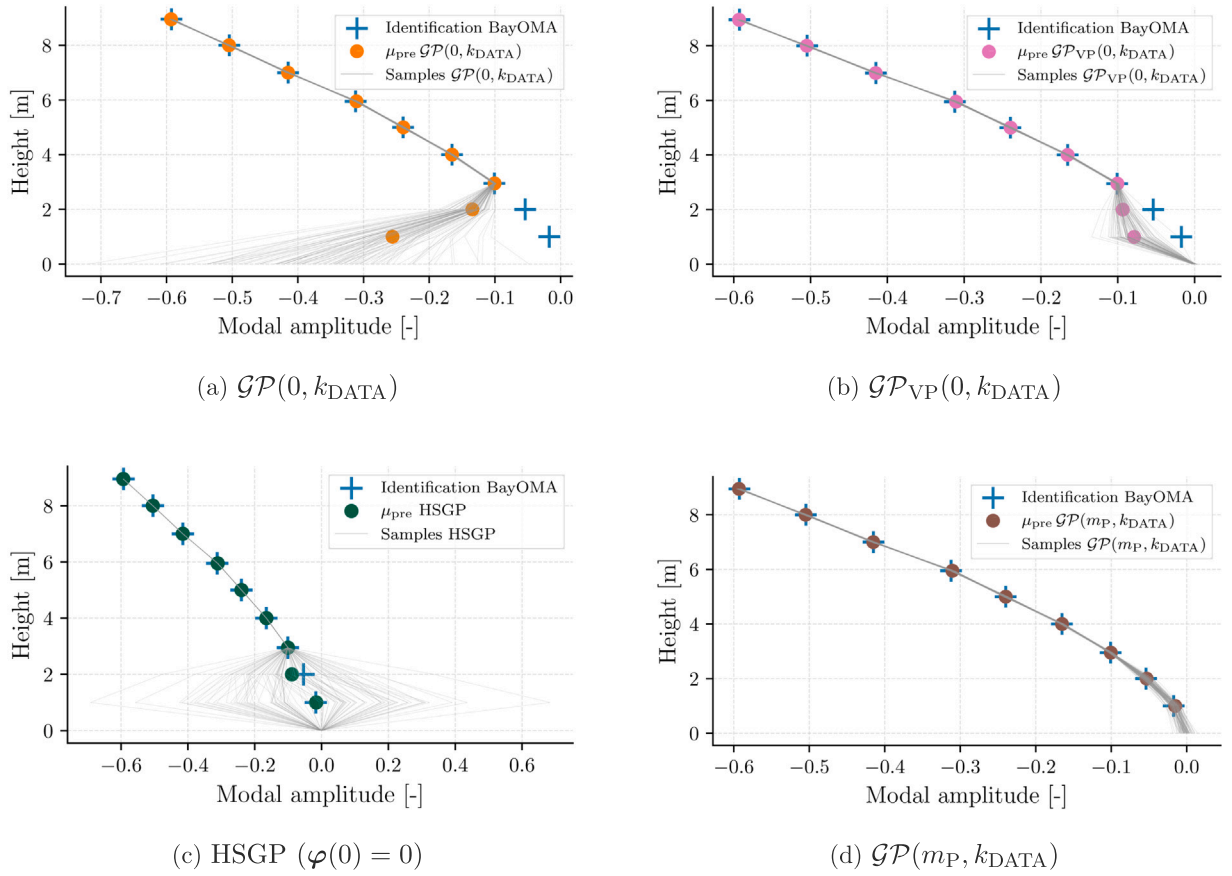


Fig. 6. Predicted mode shape of B1-x at 1.34°C for different grey-box models with uncertainty visualised by samples drawn from the joint GP posterior predictive distribution; the two lowest MLs (ML8 and ML9) are excluded from the training data: standard GP (6(a)), virtual point GP (6(b)), HSGP (6(c)), and GP with prior mean (6(d)).

Overall, when training data are highly sparse, it is preferable to impose only reliable prior knowledge. For mode shape prediction, for example, enforcing the boundary condition $\boldsymbol{\varphi}(0) = 0$ provides this kind of reliable prior knowledge. As training density increases, the predictive accuracy of the different models converges, and the data-driven component dominates while prior knowledge becomes secondary. Accordingly, any of the models considered may be employed from an SHM practitioner's perspective.

4.2. Limited number of spatial measurement levels

A further challenge for continuous SHM based on mode shapes arises when measurements from specific MLs are unavailable, either due to sensors having yet to be installed (e.g., there is a restricted budget, limiting the total number of sensors that can be used) or failure, whilst predictions at those locations are still required. To study this scenario, the predictive accuracy of the grey-box models is evaluated under limited coverage of measurement locations. For this purpose, the well-covered training set with a temperature-bin width of 0.25°C from the previous analysis scenario is used, but measurements from the two lowest levels (ML8 and ML9; see Fig. 2(b)) are excluded from the training. The predictions of the models for the entire mode shape of the first bending mode B1-x at 1.34°C (median temperature of the test data) are shown in Fig. 6. In addition, for each model, 100 samples drawn from the joint GP posterior predictive distribution are included to visualise predictive uncertainty across all MLs.

Fig. 6 illustrates that the purely data-driven model, $\mathcal{GP}(0, k_{\text{DATA}})$ (Fig. 6(a)), fails to reconstruct the full mode shape when the training data provide incomplete measurement-location coverage. Predictive uncertainty increases towards the clamped end, as these MLs are farther away from the lowest ML included in the training set. Adding the displacement constraint $\boldsymbol{\varphi}(0) = 0$ at the clamped end improves performance: both the trivial virtual-point model $\mathcal{GP}_{\text{VP}}(0, k_{\text{DATA}})$ (Fig. 6(b)) and the HSGP (Fig. 6(c)) predict more accurately in regions beyond the training range. The HSGP achieves higher accuracy than the virtual-point variant because the sine eigenbasis of the Laplacian yields eigenfunctions that approximate LUMO's bending modes [47]. For the B1-x mode shape, four representative HSGP eigenfunctions are illustrated in Fig. 7.

The four eigenfunctions shown in Fig. 7 qualitatively resemble the mode shapes of LUMO's first two bending-mode pairs. Two observations regarding predictive uncertainty are worth noting. First, for both the virtual-point model and the HSGP, predictive

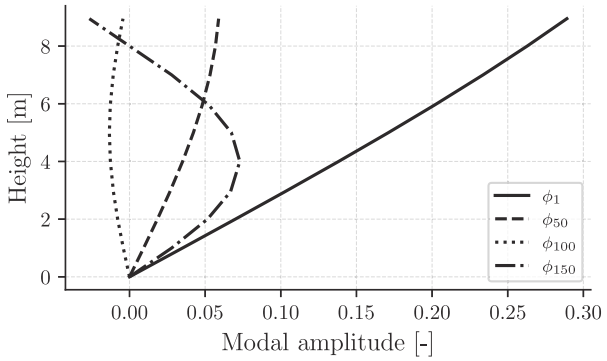


Fig. 7. Realisations of the HSGP eigenfunctions ϕ_i for predicting mode shape B1-x, illustrated for $i = 1, 50, 100, 150$.

Table 3
Comparison of MAC, nMSE, and MNLPD for the first and second bending mode shapes (B1-y, B1-x, B2-y, B2-x) predicted by different grey-box models. $\boldsymbol{\varphi}$ denotes a vector-valued quantity, whereas μ_{pre} denotes the scalar posterior mean and σ^2_{pre} the posterior variance. $\overline{\text{MAC}}$ represents the average over all predictions. In each row, the best-performing model is highlighted in bold.

Identifier	Metric	$\mathcal{GP}(0, k_{\text{DATA}})$	$\mathcal{GP}_{\text{VP}}(0, k_{\text{DATA}})$	HSGP	$\mathcal{GP}(m_{\text{P}}, k_{\text{DATA}})$
B1-y	$\overline{\text{MAC}}(\boldsymbol{\varphi}_{\text{pre}}, \boldsymbol{\varphi}_{\text{true}})$	0.963	0.997	0.997	**1.000**
	$\text{nMSE}(\mu_{\text{pre}}, y_{\text{true}})$	5.061	0.434	0.360	**0.010**
	$\text{MNLPD}(\mu_{\text{pre}}, \sigma^2_{\text{pre}}, y_{\text{true}})$	−2.012	−2.811	−3.674	**−5.374**
B1-x	$\overline{\text{MAC}}(\boldsymbol{\varphi}_{\text{pre}}, \boldsymbol{\varphi}_{\text{true}})$	0.937	0.993	0.998	**1.000**
	$\text{nMSE}(\mu_{\text{pre}}, y_{\text{true}})$	12.502	1.258	0.290	**0.007**
	$\text{MNLPD}(\mu_{\text{pre}}, \sigma^2_{\text{pre}}, y_{\text{true}})$	−3.844	−4.304	−4.818	**−5.412**
B2-y	$\overline{\text{MAC}}(\boldsymbol{\varphi}_{\text{pre}}, \boldsymbol{\varphi}_{\text{true}})$	0.956	0.994	0.996	**1.000**
	$\text{nMSE}(\mu_{\text{pre}}, y_{\text{true}})$	5.542	0.664	0.450	**0.014**
	$\text{MNLPD}(\mu_{\text{pre}}, \sigma^2_{\text{pre}}, y_{\text{true}})$	−1.177	1.221	−1.746	**−4.754**
B2-x	$\overline{\text{MAC}}(\boldsymbol{\varphi}_{\text{pre}}, \boldsymbol{\varphi}_{\text{true}})$	0.904	0.990	0.994	**0.999**
	$\text{nMSE}(\mu_{\text{pre}}, y_{\text{true}})$	11.240	1.158	0.716	**0.067**
	$\text{MNLPD}(\mu_{\text{pre}}, \sigma^2_{\text{pre}}, y_{\text{true}})$	−3.451	−3.118	−4.328	**−4.432**

uncertainty decreases towards the clamped end, as prior knowledge is introduced through the displacement constraint at that boundary. Second, the virtual-point approach exhibits lower predictive uncertainty than the HSGP at unobserved MLs, yet its posterior mean is inaccurate there, indicating overconfidence in these regions. The uncertainty estimates of the HSGP, by contrast, are more consistent with its predictive performance. The grey-box model with a physics-based prior mean, $\mathcal{GP}(m_{\text{P}}, k_{\text{DATA}})$, depicted in Fig. 6(d), achieves the best overall agreement with the true mode shape, including in unobserved regions. This behaviour is consistent with the informative structure introduced through the prior mean, which stabilises predictions beyond the observed ML range. However, compared with the virtual-point approach and the HSGP, predictive uncertainty near the clamped end is larger, since the displacement constraint is enforced only implicitly through the prior mean rather than explicitly. Assessing such differences in predictive uncertainty requires a metric that evaluates not only the posterior mean but also the quality of the predictive distribution. To this end, the mean negative log predictive density (MNLPD) is introduced as a complementary metric alongside the nMSE (cf. Eq. (12)). The MNLPD over all test points is computed as:

$$\text{MNLPD} = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{2}\log(2\pi\sigma_i^2) + \frac{(y_i - \mu_i)^2}{2\sigma_i^2}\right), \tag{13}$$

where μ_i and σ_i^2 denote the posterior predictive mean and variance at the ith test point, respectively. The first term penalises unnecessarily large predictive variances, while the second penalises inaccurate point predictions relative to the predictive uncertainty. Unlike the nMSE, which evaluates only the posterior mean, the MNLPD therefore assesses how well the full predictive distribution explains the test data. Although absolute MNLPD values are difficult to interpret in isolation, comparisons across models are meaningful: a lower MNLPD indicates a model that is both more accurate and better calibrated.

To facilitate quantitative comparison, the nMSE, MNLPD and MAC for the test data are reported in Table 3. The MAC measures the overall mode shape similarity, the nMSE evaluates pointwise discrepancies in the posterior mean, and the MNLPD additionally assesses the quality of the predictive uncertainty.

Table 3 confirms the qualitative observations in Fig. 6: increasing the amount of incorporated prior knowledge consistently improves model performance, as reflected by monotonically decreasing nMSE and MNLPD values from $\mathcal{GP}(0, k_{\text{DATA}})$ to $\mathcal{GP}(m_{\text{P}}, k_{\text{DATA}})$.

A notable exception is the virtual-point approach, which exhibits a degraded MNLPD relative to $\mathcal{GP}(0, k_{\text{DATA}})$ for the second bending mode shapes. For B2-y in particular, it yields a markedly larger MNLPD despite achieving a relatively low nMSE compared with the black-box baseline, pointing to poor calibration of the predictive distribution. This is consistent with the overconfidence observed in Fig. 6(c), where the virtual-point approach assigns overly small predictive variances despite inaccurate point predictions. Finally, the MAC values confirm that both the HSGP and $\mathcal{GP}(m_{\text{P}}, k_{\text{DATA}})$ accurately recover the full mode shapes even in the presence of missing measurement levels.

For practical SHM, when measurement-location coverage is incomplete and prior knowledge in the form of an FE mode shape is available, the grey-box model should incorporate this explicitly through the prior mean m_{P}. Where such knowledge is unavailable, the HSGP remains a sound choice for prediction under limited ML coverage. By contrast, a purely data-driven model is not advisable in this setting.

4.3. Prediction for unknown temperatures

The third analysis scenario addresses out-of-range temperatures, arguably the most challenging setting for mode shape prediction. Such incomplete EOV coverage is common when the training period is short, monitoring starts soon after construction, or rare weather regimes occur. Accordingly, the proposed grey-box models are evaluated under previously unseen temperature ranges. Specifically, for LUMO, measurements at very low temperatures are assumed to be unavailable; the training set is thus restricted to temperatures above −5 °C, reflecting the site conditions. Note that the available prior knowledge is limited to spatial information and does not encode temperature effects. Therefore, the prior mean m_{P} of $\mathcal{GP}(m_{\text{P}}, k_{\text{DATA}})$ constrains solutions to physically plausible shapes but cannot capture the nonlinear step across 0°C observed in Fig. 3(b). Consequently, this step cannot be learned without representative training data [34].

The grey-box models are tested on data below −5 °C, requiring them to make predictions for temperatures beyond their training range. Fig. 8 shows an example prediction for mode shape B1-x at the lowest temperature (−9 °C) included in the test data. In addition, for each model, 100 samples drawn from the joint GP posterior predictive distribution are included to visualise predictive uncertainty across all MLs.

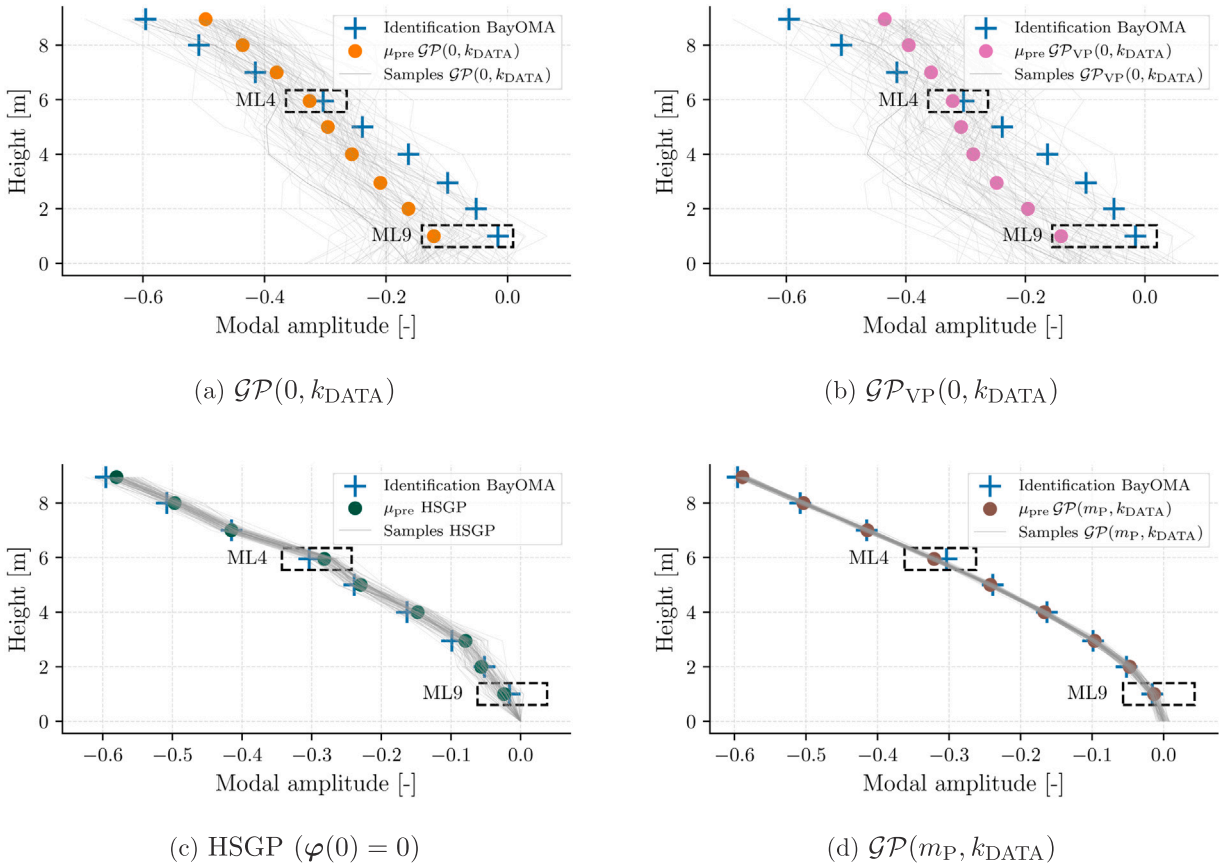
Both the black-box model $\mathcal{GP}(0, k_{\text{DATA}})$ and the grey-box model $\mathcal{GP}_{\text{VP}}(0, k_{\text{DATA}})$ (Figs. 8(a) and 8(b)) fail to reproduce the B1-x mode shape under previously unseen temperature conditions. In both cases, predictive uncertainty remains large, indicating low confidence in the corresponding predictions, and the prior knowledge introduced via virtual points does not appear to reduce it. In contrast, the HSGP (Fig. 8(c)) delivers substantially better predictions. Although it enforces only the displacement condition $\boldsymbol{\varphi}(0) = 0$ at the clamped end, with covariance represented via a Laplacian sine-eigenfunction expansion with Dirichlet boundary conditions (see also Section 4.2), the incorporated prior knowledge is purely spatial and limited to this boundary constraint. This is consistent with the reduced prediction accuracy away from the clamped end, most notably at LUMO's free end (ML1), with additional discrepancies at ML4 (highlighted in Fig. 8(c)) and ML7. Regarding predictive uncertainty, it is also evident that uncertainty across the higher MLs is larger than in the previous analysis scenario (cf. Fig. 6(c)). This is expected: in the present scenario, the unobserved variable is temperature, which affects predictions at all MLs, whereas previously the missing information was confined to the lowest two MLs. Importantly, uncertainty near the clamped end remains comparatively low in Fig. 8(c), indicating that explicitly enforcing $\boldsymbol{\varphi}(0) = 0$ locally reduces predictive uncertainty even when temperature is unobserved.

The grey-box model $\mathcal{GP}(m_{\text{P}}, k_{\text{DATA}})$, shown in Fig. 8(d), again achieves the highest overall accuracy under unobserved temperatures and exhibits the lowest predictive uncertainty among the models in Fig. 8, although this uncertainty is higher than in the previous analysis scenario (cf. Fig. 6(d)). These results show that explicitly encoding prior knowledge via the GP prior mean substantially reduces predictive uncertainty. While this is beneficial when the prior mean is informative, this can also lead to overconfident predictions if the prior mean is misspecified, since the model may remain highly confident even when the physics-based prior mean is misleading.

To examine the ML-wise behaviour in more detail, particularly in regions where boundary effects are expected to be most influential, Fig. 9 shows predictions across the full temperature range for ML9, the measurement level closest to the clamped end.

At previously unseen temperatures, neither the $\mathcal{GP}(0, k_{\text{DATA}})$ nor the $\mathcal{GP}_{\text{VP}}(0, k_{\text{DATA}})$ can predict the modal amplitude at ML9; see Fig. 9(a). The poor predictive accuracy of $\mathcal{GP}(0, k_{\text{DATA}})$ is expected, as a purely data-driven GP cannot predict beyond its training range. The introduction of virtual point constraints also fails to improve the robustness of the model: the added boundary condition only has a local influence, and since ML height and temperature form a two-dimensional input surface, the boundary condition $\boldsymbol{\varphi}(0) = 0$ could only be approximated (see Section 2). At a temperature of −9 °C, the model predicts $\boldsymbol{\varphi}(0) = 0.1123$, indicating that the available training samples from state StE are insufficient to enforce the boundary constraint accurately. By contrast, Fig. 9(b) shows that, for out-of-range temperatures, $\mathcal{GP}(m_{\text{P}}, k_{\text{DATA}})$ reverts to its prior mean m_{P}, thereby limiting the prediction error relative to the models in Fig. 9(a); note the markedly different y-axis scales in Figs. 9(a) and 9(b). The HSGP also performs comparatively well near the clamped end owing to its sine eigenbasis of the Laplacian. The squared-error results in Fig. 10 confirm that the HSGP and $\mathcal{GP}(m_{\text{P}}, k_{\text{DATA}})$ substantially reduce the error relative to the other approaches, whereas the virtual-point model even exceeds the error of the black-box baseline.

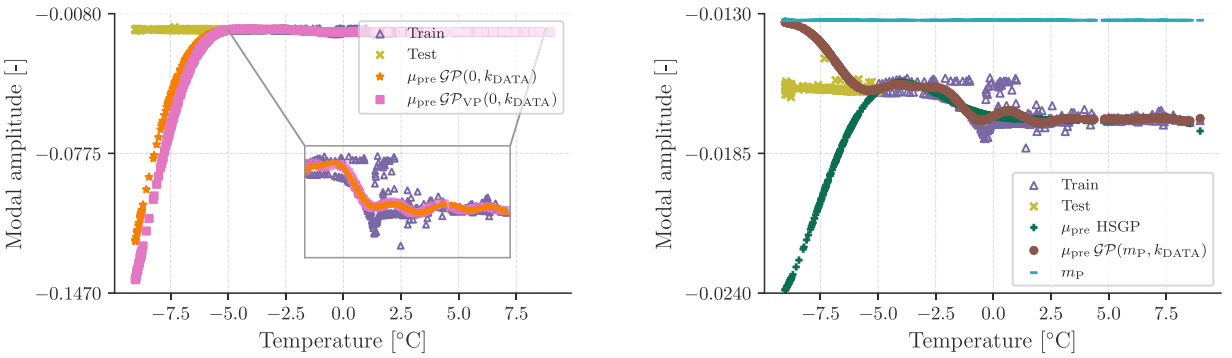
After examining the predictions near the clamped end, where the prior knowledge $\boldsymbol{\varphi}(0) = 0$ is most relevant, the next step is to investigate predictive accuracy at higher MLs. Since Fig. 8(d) shows that the most significant deviation for $\mathcal{GP}(m_{\text{P}}, k_{\text{DATA}})$ occurs at ML4, Fig. 11 presents the temperature-dependent predictions at ML4 for the black-box GP, the HSGP, and $\mathcal{GP}(m_{\text{P}}, k_{\text{DATA}})$ across the full temperature range of State StE. Since $\mathcal{GP}_{\text{VP}}(0, k_{\text{DATA}})$ did not offer an improvement over the black-box model, this approach is omitted from this analysis.



(a) $\mathcal{GP}(0, k_{\text{DATA}})$ (b) $\mathcal{GP}_{\text{VP}}(0, k_{\text{DATA}})$

(c) HSGP $(\varphi(0) = 0)$ (d) $\mathcal{GP}(m_{\text{P}}, k_{\text{DATA}})$

Fig. 8. Mode shape B1-x predictions of the GP models at an unseen temperature of −9 °C, including samples drawn from the joint GP posterior predictive distribution: standard GP (8(a)), virtual point GP (8(b)), HSGP (8(c)), and GP with prior mean (8(d)). ML9 and ML4 are highlighted as ML9 is closest to the clamped end, and $\mathcal{GP}(m_{\text{P}}, k_{\text{DATA}})$ exhibits a noticeable deviation at ML4.



(a) $\mathcal{GP}(0, k_{\text{DATA}})$ and $\mathcal{GP}_{\text{VP}}(0, k_{\text{DATA}})$ (b) HSGP and $\mathcal{GP}(m_{\text{P}}, k_{\text{DATA}})$

Fig. 9. Predictions of the black-box model and the virtual point approach for the modal amplitude at ML9 across the full temperature range for mode shape B1-x (9(a)). Fig. 9(b) contrasts the HSGP and prior-mean GP predictions. Note the different y-axis scales in Figs. 9(a) and 9(b).

Within the training range, the black-box GP, the HSGP, and the prior-mean grey-box model $\mathcal{GP}(m_{\text{P}}, k_{\text{DATA}})$ all reproduce the measured temperature-dependent trend. Beyond the training data, the predictions of the considered models revert, by construction, to their respective prior means. For $\mathcal{GP}(m_{\text{P}}, k_{\text{DATA}})$, this means falling back to the FE-based prior mean m_{P}, so its extrapolations remain physically consistent with the FE mode shapes. Because the HSGP operates on the original targets, its predictions at temperatures not represented in the training data differ from those of the black-box model.

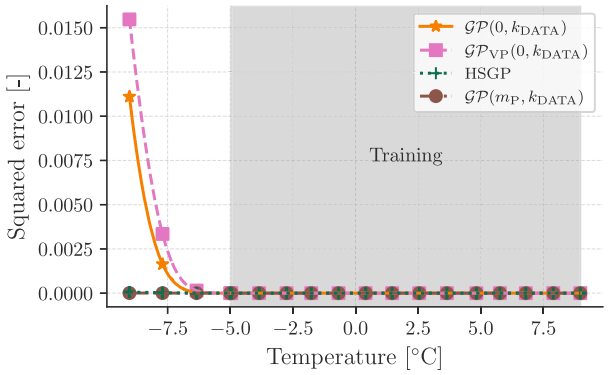


Fig. 10. Squared error of all models for mode shape B1-x at ML9.

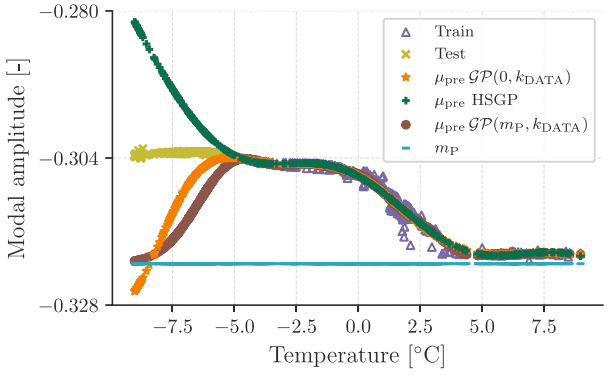


Fig. 11. Results for the prediction of the modal amplitude at ML4 over the complete temperature range for B1-x.

Table 4
Comparison of MAC, nMSE, and MNLPD for the first and second bending mode shapes (B1-y, B1-x, B2-y, B2-x) predicted by different grey-box models. $\boldsymbol{\varphi}$ denotes a vector-valued quantity, whereas μ_{pre} denotes the scalar posterior mean and σ^2_{pre} the posterior variance. $\overline{\mathrm{MAC}}$ represents the average over all predictions. In each row, the best-performing model is highlighted in bold.

Identifier	Metric	$\mathcal{GP}(0, k_{\mathrm{DATA}})$	$\mathcal{GP}_{\mathrm{VP}}(0, k_{\mathrm{DATA}})$	HSGP	$\mathcal{GP}(m_{\mathrm{P}}, k_{\mathrm{DATA}})$
B1-y	$\overline{\mathrm{MAC}}(\boldsymbol{\varphi}_{\mathrm{pre}}, \boldsymbol{\varphi}_{\mathrm{true}})$	0.999	0.986	0.999	**1.000**
	$\mathrm{nMSE}(\mu_{\mathrm{pre}}, y_{\mathrm{true}})$	0.179	1.821	0.092	**0.029**
	$\mathrm{MNLPD}(\mu_{\mathrm{pre}}, \sigma^2_{\mathrm{pre}}, y_{\mathrm{true}})$	−3.879	−3.325	**−4.661**	−4.564
B1-x	$\overline{\mathrm{MAC}}(\boldsymbol{\varphi}_{\mathrm{pre}}, \boldsymbol{\varphi}_{\mathrm{true}})$	0.990	0.977	1.000	**1.000**
	$\mathrm{nMSE}(\mu_{\mathrm{pre}}, y_{\mathrm{true}})$	1.802	4.060	0.140	**0.041**
	$\mathrm{MNLPD}(\mu_{\mathrm{pre}}, \sigma^2_{\mathrm{pre}}, y_{\mathrm{true}})$	−3.622	−3.116	**−4.621**	−4.353
B2-y	$\overline{\mathrm{MAC}}(\boldsymbol{\varphi}_{\mathrm{pre}}, \boldsymbol{\varphi}_{\mathrm{true}})$	0.800	0.984	0.999	**0.999**
	$\mathrm{nMSE}(\mu_{\mathrm{pre}}, y_{\mathrm{true}})$	24.101	1.812	0.140	**0.120**
	$\mathrm{MNLPD}(\mu_{\mathrm{pre}}, \sigma^2_{\mathrm{pre}}, y_{\mathrm{true}})$	−2.458	−3.430	−3.427	**−3.711**
B2-x	$\overline{\mathrm{MAC}}(\boldsymbol{\varphi}_{\mathrm{pre}}, \boldsymbol{\varphi}_{\mathrm{true}})$	0.974	0.968	0.999	**0.999**
	$\mathrm{nMSE}(\mu_{\mathrm{pre}}, y_{\mathrm{true}})$	2.936	3.801	0.155	**0.078**
	$\mathrm{MNLPD}(\mu_{\mathrm{pre}}, \sigma^2_{\mathrm{pre}}, y_{\mathrm{true}})$	−3.363	−2.990	−3.762	**−3.945**

Table 4 summarises the quantitative results for mode shapes B1-y, B1-x, B2-y, and B2-x, reporting MAC, nMSE, and MNLPD values evaluated on the test set.

As shown in Table 4, incorporating prior knowledge through the grey-box models HSGP or $\mathcal{GP}(m_{\mathrm{P}}, k_{\mathrm{DATA}})$ improves prediction accuracy at unseen temperatures compared with both the trivial virtual-point model $\mathcal{GP}_{\mathrm{VP}}(0, k_{\mathrm{DATA}})$ and the black-box baseline. For the first bending-mode pair, the HSGP achieves lower MNLPD values than $\mathcal{GP}(m_{\mathrm{P}}, k_{\mathrm{DATA}})$, even though the latter yields lower nMSE values. This discrepancy suggests that $\mathcal{GP}(m_{\mathrm{P}}, k_{\mathrm{DATA}})$ produces overly small predictive variances in the extrapolation regime, resulting in overconfident predictions despite accurate posterior means. It is also worth noting that, although the black-box baseline

Table 5
Characterisation of the reference states. The four mode shapes (B1-x, B1-y, B2-x, B2-y) belonging to each state (Reference 1, Reference 2, Reference 3) were identified using BayOMA.

Name	Temperature [°C]	Date
Reference 1	15.75	21-Apr-2021 07:39
Reference 2	5.26	21-Apr-2021 19:59
Reference 3	19.53	21-Apr-2021 08:59

and the virtual-point approach exhibit larger nMSE values, their MNLPD values are not penalised as heavily as those nMSE values might suggest. This indicates that their broader predictive distributions partially compensate for inaccurate posterior means under out-of-range conditions.

Overall, for temperatures outside the training range, the HSGP is particularly effective near the clamped end, where the enforced zero-displacement boundary condition provides a firm guide for mode-shape prediction. This advantage does not necessarily carry over to higher MLs. Accordingly, if the prediction location is unspecified or spans higher MLs, $\mathcal{GP}(m_{\mathrm{P}}, k_{\mathrm{DATA}})$ is a more robust choice from a SHM practitioner's perspective, as it delivers more uniform performance across the structure.

5. Application for damage localisation via model updating

Finally, a prescribed damage scenario at LUMO is considered to assess continuous, model-based damage localisation using mode shapes alone, thereby isolating the impact of mode shape data normalisation. When employing the relative least-squares error metric, or any other discrepancy metric, reliable localisation is linked to accurate mode shape predictions across the range of EOVs, which, in turn, is crucial for reasonable maintenance decision-making. As in the third analysis scenario (cf. Section 4.3), SHM operators typically face limited EOV coverage in the available training data for continuous monitoring, which is arguably the most challenging setting for mode shape prediction. The final analysis scenario, therefore, evaluates how the proposed methods affect localisation accuracy when EOV coverage in the training data is limited. In practical SHM, this means starting monitoring soon after installation, whilst training data do not span the EOV range expected under damage. Generally, grey-box models can eliminate the need to select a representative reference shape manually, and unlike black-box models, yield reliable mode shape predictions even for EOV regimes that are poorly represented in the training data.

For this analysis scenario, state H serves as the undamaged dataset and only temperatures above 15°C are included in the training data. Damage localisation is assessed for a single strut removal at location DAM6 (Table 1). Note that our aim is not to exhaustively localise damage at the LUMO. Rather, the analysis examines how mode shapes can be predicted under different kinds of limited training data and how these predictions can be leveraged for model-based damage localisation. For a comprehensive treatment of model-based damage localisation, the interested reader is referred to Ragnitz et al. [8].

5.1. Damage localisation without data normalisation

To initially demonstrate the advantage of regression-based data normalisation within the relative metric approach, three reference states for $\boldsymbol{\varphi}_{M_0,k}$, $k = 1, \ldots, N_{\mathrm{modes}}$, have been selected; these are summarised in Table 5.

The authors acknowledge that the choice of the reference states listed in Table 5 could be improved, for example, by applying clustering techniques such as those proposed by Ragnitz et al. [23]. However, even with this approach, the task of defining clusters – and thus choosing an appropriate reference state – remains. Fig. 12 presents the localisation results for the damage scenario DAM6 using the reference states listed in Table 5.

Fig. 12 shows that using the reference states of Table 5 in Eq. (11) makes the result highly sensitive to which reference is chosen. When the chosen reference is good (e.g., corresponds to the mean temperature of state DAM6 (15.33 °C)), the correct damage position is identified in 94.76% of the 725 datasets examined (Reference 1). A less suitable reference reduces this rate to 66.21% (Reference 2), and in the worst case, damage localisation succeeds in only 8.69% of the datasets (Reference 3). As it is challenging to select an appropriate reference mode *a priori*, and potential misidentifications due to an unsuitable reference mode can be costly – particularly for hard-to-access structures such as offshore wind turbines – it is reasonable to model the reference mode shape as an explicit function of EOVs via a regression-based data-normalisation scheme, thus avoiding manual selection of the reference mode.

5.2. Damage localisation with data normalisation

To assess the impact of the proposed methods on localisation accuracy under limited EOV coverage in the training data, two training regimes are considered: (i) all data from State H are used for training; the mean temperatures are similar for State H (11.31 °C) and the damaged state DAM6 (15.33 °C); (ii) to emulate early post-installation SHM, the training set is restricted to temperatures above 15 °C. In the latter, temperatures below the damaged state's mean are absent, corresponding to initiating monitoring shortly after installation, before the training data span the full EOV range present in the damaged state. To integrate regression-based data normalisation into the model-updating procedure for damage localisation, an appropriate grey-box model

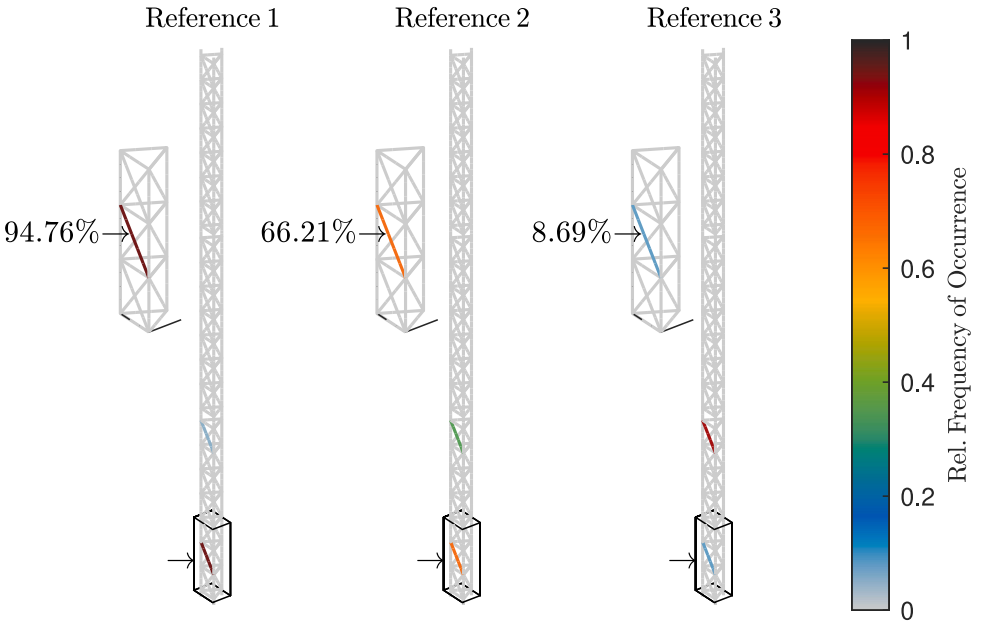


Fig. 12. Results of the damage localisation using three different fixed reference mode shapes. The arrow indicates the true position of the removed strut.

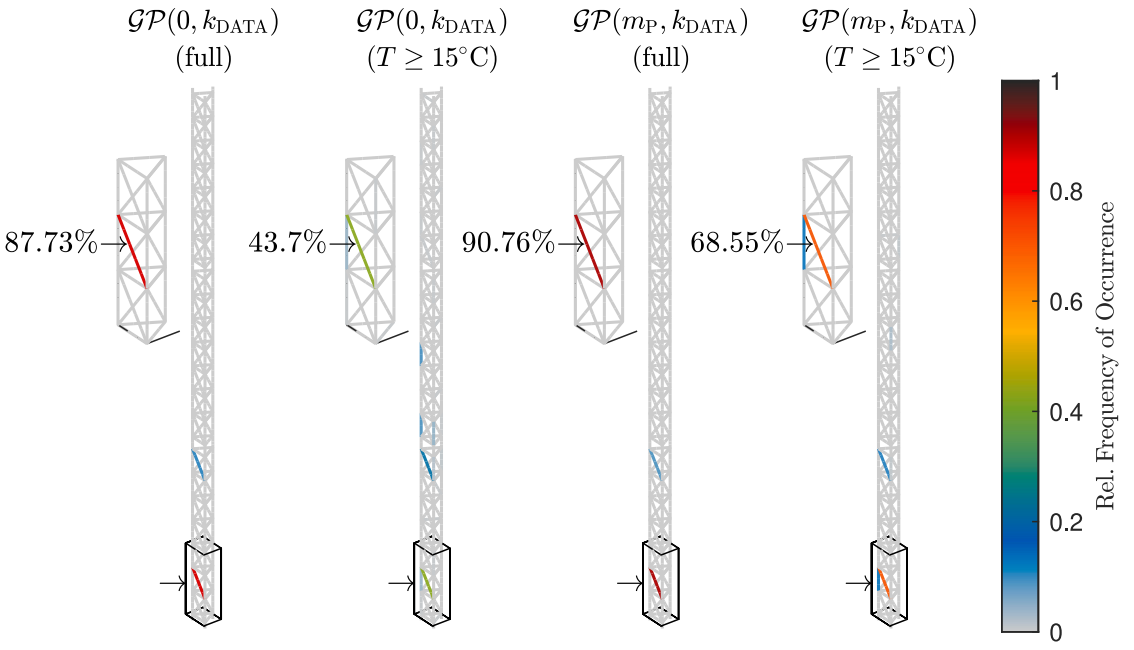


Fig. 13. Results of the damage localisation for full and incomplete training data ($T \geq 15$ °C) using black-box and grey-box-based data normalisation schemes. The arrow indicates the true position of the removed strut, and T represents the temperature.

must be selected. As stated in the third analysis scenario (Section 4.3), $\mathcal{GP}(m_{\mathrm{P}}, k_{\mathrm{DATA}})$ is preferred to the HSGP for predictions at temperatures outside the training range, unless prior expert knowledge gives reasons to focus on the area near the clamped end. Because no such assumption is generally justified for LUMO, the present analysis employs the grey-box model with the FE-based prior mean m_{P}. Throughout this analysis scenario, this grey-box model is benchmarked against the black-box baseline $\mathcal{GP}(0, k_{\mathrm{DATA}})$. Both models are applied to model the reference mode shapes $\boldsymbol{\varphi}_{M_0,k}$, $k = 1, \ldots, N_{\mathrm{modes}}$ (cf. Eq. (11)), with respect to temperature variations. The results for the damage localisation of state DAM6 are depicted in Fig. 13.

Applying both data normalisation strategies within the model updating yields good localisation accuracy when using complete training data. Here, the black-box model $\mathcal{GP}(0, k_{\mathrm{DATA}})$ identifies the damage location in 87.73% of the 725 evaluated datasets. In comparison, the grey-box model correctly identifies the damage position in 90.76% of the evaluated datasets. This result is consistent with the assumption that, as the training set becomes more complete, the grey- and black-box models exhibit comparable accuracy, since the data-driven component dominates in both. Consequently, the grey-box model $\mathcal{GP}(m_{\mathrm{P}}, k_{\mathrm{DATA}})$ can learn the discrepancies between the FE model and the measurements, making it superior to simply adopting the FE mode shape as a fixed reference. When the training data are incomplete (temperature $\geq 15°\mathrm{C}$), the black-box model identifies the correct damage position in only 43.70% of the datasets, while the grey-box model achieves 68.55%. The grey-box model identifies the damage position more accurately because, in the absence of observations, it reverts to its physics-based prior mean. Consequently, in the relative metric (cf. Eq. (11)), $\boldsymbol{\varphi}_{S_0,k} = \boldsymbol{\varphi}_{M_0,k}$, so the corresponding terms cancel, and zero-order errors may arise. Model updating then depends on the quality of the simulated data for the unknown temperatures.

6. Conclusions and future work

This paper proposes a regression-based data normalisation scheme that models mode shapes as functions of spatial coordinate and EOVs. It embeds this representation into a relative least-squares objective for model-based SHM. To further address the issue of limited training data, two GP-based grey-box approaches are compared that incorporate physical knowledge in two ways. First, constraint-based models impose the clamped end displacement condition $\boldsymbol{\varphi}(0) = 0$ either through an HSGP or, more trivially, through virtual points. Second, a physics-based prior-mean GP uses FE mode shapes as the prior mean m_{P}, leaving EOV effects to the data-driven residual.

Using the Leibniz University Test Structure for Monitoring (LUMO), three analysis scenarios are evaluated that reflect practical SHM challenges: sparse training data, limited measurement locations, and out-of-range temperatures. Across these settings, grey-box models that embed engineering knowledge consistently outperform a black-box baseline. A prior mean GP is appealing when predictions are required across the entire structure, and some EOV regimes are missing from the training data, whereas the HSGP is preferable for highly sparse data or for out-of-range predictions near the clamped end. The virtual-point approach $\mathcal{GP}_{\mathrm{VP}}(0, k_{\mathrm{DATA}})$ proves unsuitable under limited training data. Finally, the grey-box normalisation is integrated into continuous, model-based damage localisation on LUMO using mode shapes alone: using the prior-mean GP, the true damage location is identified in 90.76% of datasets with full training coverage and in 68.55% when training is restricted to $T \geq 15\ °\mathrm{C}$ (versus 43.70% for the black-box GP). These results provide clear guidance: impose boundary-aware HSGP under severe sparsity or near the clamp; otherwise, prefer an FE-informed prior-mean GP for robust, real-time-capable model-based damage localisation.

This work provides a foundation for grey-box-enhanced model updating in long-term, real-time-capable SHM. While the present study focuses on a lattice tower (LUMO), the proposed framework is general and targets model-based damage localisation under EOVs using mode shapes. Provided that practitioners can extract sufficiently meaningful mode shapes, define a suitable damage parametrisation for model-based damage localisation, and account for the relevant EOV–DSF relationships, this methodology offers broad potential across a range of SHM applications. This is especially true where prior knowledge, such as an FE model or boundary-condition information, is available. However, applying this framework to other domains, such as bridges or wind turbines, will require specific considerations unique to those structures, which will be the focus of future research. Another important direction for further work is to investigate how predictive uncertainty can be used to account for mode-shape identification uncertainty, including whether parts of this uncertainty are input-dependent and may therefore be better captured by heteroscedastic GP models. In addition, regime-switching phenomena such as freeze-thaw behaviour could be addressed using local models, such as change-point kernels or mixture-of-experts, that encode regime-specific physical priors, thereby combining different kinds of prior knowledge about the temperature dependence of the modal amplitudes.

CRediT authorship contribution statement

Sören Möller: Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Conceptualization. **Matthew R. Jones:** Writing – review & editing, Methodology, Investigation, Formal analysis, Conceptualization. **Clemens Jonscher:** Writing – review & editing, Methodology, Formal analysis. **Jasper Ragnitz:** Visualization, Software. **Elizabeth J. Cross:** Writing – review & editing, Supervision, Investigation, Formal analysis. **Raimund Rolfes:** Writing – review & editing, Supervision, Resources, Project administration, Funding acquisition.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the authors used "ChatGPT" and "Grammarly" in order to improve the orthography and readability. After using these tools, the authors reviewed and edited the content as needed and took full responsibility for the content of the published article.

Funding

S.M. gratefully acknowledges the financial support provided by the Deutsche Forschungsgemeinschaft, Germany (DFG, German Research Foundation) - SFB 1463 - 434502799, subproject C02. E.J.C. and M.R.J. would like to acknowledge the support of Innovate UK, United Kingdom through the OLLGA project grant 10040817. M.R.J. would like to acknowledge the University of Sheffield, United Kingdom for support through a Research Excellence Fellowship. C.J. greatly acknowledges the financial support of the Lower Saxony Ministry of Science and Culture and the Volkswagen Stiftung (research project Reallabor 70GW Offshore Wind, FKZ ZN4468).

Declaration of competing interest

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

Data availability

The data that support the findings of this article are openly available via the institutional repository of Leibniz Universität Hannover [59].

References

[1] Keith Worden, Structural health monitoring using pattern recognition, in: New Trends in Vibration Based Structural Health Monitoring, Springer Vienna, Vienna, 2010, pp. 183–246, http://dx.doi.org/10.1007/978-3-7091-0399-9_5.

[2] Xin Zhou, Chul-Woo Kim, Feng-Liang Zhang, Kai-Chun Chang, Vibration-based Bayesian model updating of an actual steel truss bridge subjected to incremental damage, Eng. Struct. 260 (2022) 114226, http://dx.doi.org/10.1016/j.engstruct.2022.114226.

[3] J.E. Mottershead, M.I. Friswell, Model updating in structural dynamics: A survey, J. Sound Vib. 167 (2) (1993) 347–375, http://dx.doi.org/10.1006/jsvi.1993.1340.

[4] Wei Fan, Pizhong Qiao, Vibration-based damage identification methods: A review and comparative study, Struct. Health Monit. 10 (1) (2011) 83–111, http://dx.doi.org/10.1177/1475921710365419.

[5] Suzana Ereiz, Ivan Duvnjak, Javier Fernando Jiménez-Alonso, Review of finite element model updating methods for structural applications, Structures 41 (2022) 684–723, http://dx.doi.org/10.1016/j.istruc.2022.05.041.

[6] Christian Günther, Benedikt Hofmeister, Clemens Hübler, Clemens Jonscher, Jasper Ragnitz, Jenny Schubert, Marc C. Steinbach, Damage location in mechanical structures by multi-objective pattern search, Optim. Eng. (2025) http://dx.doi.org/10.1007/s11081-024-09940-1.

[7] Marlene Wolniak, Benedikt Hofmeister, Clemens Jonscher, Matthias Fankhänel, Ansgar Loose, Clemens Hübler, Raimund Rolfes, Validation of an FE model updating procedure for damage assessment using a modular laboratory experiment with a reversible damage mechanism, J. Civ. Struct. Health Monit. (2023) http://dx.doi.org/10.1007/s13349-023-00701-9.

[8] Jasper Ragnitz, Benedikt Hofmeister, Clemens Jonscher, Clemens Hübler, Raimund Rolfes, A stochastic multi-objective optimisation approach for damage localisation via model updating with uncertain input parameters, Eng. Struct. 330 (2025) 119860, http://dx.doi.org/10.1016/j.engstruct.2025.119860.

[9] O.S. Salawu, Detection of structural damage through changes in frequency: a review, Eng. Struct. 19 (9) (1997) 718–723, http://dx.doi.org/10.1016/S0141-0296(96)00149-6.

[10] Charles R Farrar, Phillip J Cornwell, Scott W Doebling, Michael B Prime, Structural health monitoring studies of the alamosa canyon and I-40 bridges, 2000, http://dx.doi.org/10.2172/766805.

[11] Bart Peeters, Guido De Roeck, One-year monitoring of the Z24-bridge: environmental effects versus damage events, Earthq. Eng. Struct. Dyn. 30 (2) (2001) 149–171, http://dx.doi.org/10.1002/1096-9845(200102)30:2<149::AID-EQE1>3.0.CO;2-Z.

[12] Filippo Ubertini, Gabriele Comanducci, Nicola Cavalagli, Anna Laura Pisello, Annibale Luigi Materazzi, Franco Cotana, Environmental effects on natural frequencies of the san pietro bell tower in Perugia, Italy, and their removal for structural performance assessment, Mech. Syst. Signal Process. 82 (2017) 307–322, http://dx.doi.org/10.1016/j.ymssp.2016.05.025.

[13] Wout Weijtjens, Tim Verbelen, Gert De Sitter, Christof Devriendt, Foundation structural health monitoring of an offshore wind turbine—a full-scale case study, Struct. Health Monit. 15 (4) (2016) 389–402, http://dx.doi.org/10.1177/1475921715586624.

[14] Hui Li, Shunlong Li, Jinping Ou, Hongwei Li, Modal identification of bridges under varying environmental conditions: Temperature and wind effects, Struct. Control. Health Monit. 17 (5) (2010) 495–512, http://dx.doi.org/10.1002/stc.319.

[15] Y.Q. Ni, K.Q. Fan, G. Zheng, J.M. Ko, Automatic modal identification and variability in measured modal vectors of a cable-stayed bridge, Struct. Eng. Mech. 19 (2) (2005) 123–139, http://dx.doi.org/10.12989/sem.2005.19.2.123.

[16] Yong Xia, Hong Hao, Giovanna Zanardo, Andrew Deeks, Long term vibration monitoring of an RC slab: Temperature and humidity effect, Eng. Struct. 28 (3) (2006) 441–452, http://dx.doi.org/10.1016/j.engstruct.2005.09.001.

[17] Yong Xia, Bo Chen, Shun Weng, Yi-Qing Ni, You-Lin Xu, Temperature effect on vibration properties of civil structures: A literature review and case studies, J. Civ. Struct. Health Monit. 2 (1) (2012) 29–46, http://dx.doi.org/10.1007/s13349-011-0015-7.

[18] Clemens Jonscher, Sören Möller, Leon Liesecke, Daniel Schuster, Benedikt Hofmeister, Tanja Grießmann, Raimund Rolfes, Identification uncertainties of bending modes of an onshore wind turbine for vibration-based monitoring, Struct. Control. Health Monit. 2024 (2024) http://dx.doi.org/10.1155/2024/3280697.

[19] Stefan Wernitz, Benedikt Hofmeister, Clemens Jonscher, Tanja Grießmann, Raimund Rolfes, A new open-database benchmark structure for vibration-based structural health monitoring, Struct. Control. Health Monit. 29 (11) (2022) e3077, http://dx.doi.org/10.1002/stc.3077.

[20] C.R. Farrar, N. Dervilis, K. Worden, The past, present and future of structural health monitoring: An overview of three ages, Strain 61 (1) (2025) e12495, http://dx.doi.org/10.1111/str.12495, e12495 5547601.

[21] Babak Moaveni, Iman Behmanesh, Effects of changing ambient temperature on finite element model updating of the dowling hall footbridge, Eng. Struct. 43 (2012) 58–68, http://dx.doi.org/10.1016/j.engstruct.2012.05.009.

[22] Sebastian Schommer, Viet Ha Nguyen, Stefan Maas, Arno Zürbes, Model updating for structural health monitoring using static and dynamic measurements, Procedia Eng. 199 (2017) 2146–2153, http://dx.doi.org/10.1016/j.proeng.2017.09.156.

[23] Jasper Ragnitz, Benedikt Hofmeister, Clemens Jonscher, Clemens Hübler, Raimund Rolfes, Multi-objective model updating of tower structures featuring closely spaced modes, E-Journal Nondestruct. Test. 29 (7) (2024) http://dx.doi.org/10.58286/29663.

[24] Patrick Simon, Ronald Schneider, Matthias Baeßler, Guido Morgenthal, A Bayesian probabilistic framework for building models for structural health monitoring of structures subject to environmental variability, Struct. Control. Health Monit. 2024 (1) (2024) http://dx.doi.org/10.1155/2024/4204316.

[25] Shanglian Zhou, Wei Song, Environmental-effects-embedded model updating method considering environmental impacts, Struct. Control. Health Monit. 25 (3) (2018) e2116, http://dx.doi.org/10.1002/stc.2116.

[26] Lanxin Luo, Mingming Song, Huaqiang Zhong, Tiantao He, Limin Sun, Hierarchical Bayesian model updating of a long-span arch bridge considering temperature and traffic loads, Mech. Syst. Signal Process. 210 (2024) 111152, http://dx.doi.org/10.1016/j.ymssp.2024.111152.

[27] Clemens Jonscher, Stefan Warnken, Jasper Ragnitz, Jan Heinemeyer, David Märtins, Sören Möller, Benedikt Hofmeister, Raimund Rolfes, Real-time capable model updating of digital twins: A case study of a cantilever beam, in: Alvaro Cunha, Elsa Caetano (Eds.), Experimental Vibration Analysis for Civil Engineering Structures, Springer Nature Switzerland, Cham, 2025, pp. 1–10, http://dx.doi.org/10.1007/978-3-031-96110-6_99.

[28] K. Worden, E.J. Cross, On switching response surface models, with applications to the structural health monitoring of bridges, Mech. Syst. Signal Process. 98 (2018) 139–156, http://dx.doi.org/10.1016/j.ymssp.2017.04.022.

[29] Callum Roberts, David Garcia Cava, Luis D Avendaño-Valencia, Addressing practicalities in multivariate nonlinear regression for mitigating environmental and operational variations, Struct. Health Monit. 22 (2) (2023) 1237–1255, http://dx.doi.org/10.1177/14759217221091907.

[30] Clemens Jonscher, Sören Möller, Leon Liesecke, Benedikt Hofmeister, Tanja Grießmann, Raimund Rolfes, Heteroscedastic Gaussian processes for data normalisation in probabilistic novelty detection of a wind turbine, Eng. Struct. 305 (2024) 117786, http://dx.doi.org/10.1016/j.engstruct.2024.117786.

[31] Laura von Rueden, Sebastian Mayer, Katharina Beckh, Bogdan Georgiev, Sven Giesselbach, Raoul Heese, Birgit Kirsch, Michal Walczak, Julius Pfrommer, Annika Pick, Rajkumar Ramamurthy, Jochen Garcke, Christian Bauckhage, Jannis Schuecker, Informed machine learning - A taxonomy and survey of integrating prior knowledge into learning systems, IEEE Trans. Knowl. Data Eng. (2021) 1, http://dx.doi.org/10.1109/TKDE.2021.3079836.

[32] Marcus Haywood-Alexander, Wei Liu, Kiran Bacsa, Zhilu Lai, Eleni Chatzi, Discussing the spectrum of physics-enhanced machine learning: a survey on structural mechanics applications, Data-Centric Eng. 5 (2024) e31, http://dx.doi.org/10.1017/dce.2024.33.

[33] Alice Cicirello, Physics-enhanced machine learning: a position paper for dynamical systems investigations, J. Phys.: Conf. Ser. 2909 (1) (2024) 012034, http://dx.doi.org/10.1088/1742-6596/2909/1/012034.

[34] Elizabeth J. Cross, Timothy J. Rogers, Daniel J. Pitchforth, Samuel J. Gibson, Sikai Zhang, Matthew R. Jones, A spectrum of physics-informed Gaussian processes for regression in engineering, Data-Centric Eng. 5 (2024) e8, http://dx.doi.org/10.1017/dce.2024.2.

[35] Carl Edward Rasmussen, Christopher K.I. Williams, Gaussian Processes for Machine Learning, The MIT Press, ISBN: 9780262256834, 2005, http://dx.doi.org/10.7551/mitpress/3206.001.0001.

[36] Sikai Zhang, Elizabeth J. Cross, Grey-box modelling via Gaussian process mean functions for mechanical systems, in: Ramin Madarshahian, Francois Hemez (Eds.), Data Science in Engineering, Volume 9, Springer International Publishing, Cham, 2022, pp. 161–168, http://dx.doi.org/10.1007/978-3-030-76004-5_19.

[37] D.J. Pitchforth, T.J. Rogers, U.T. Tygesen, E.J. Cross, Grey-box models for wave loading prediction, Mech. Syst. Signal Process. 159 (2021) 107741, http://dx.doi.org/10.1016/j.ymssp.2021.107741.

[38] Elizabeth J. Cross, Timothy J. Rogers, Physics-derived covariance functions for machine learning in structural dynamics, IFAC-PapersOnLine 54 (7) (2021) 168–173, http://dx.doi.org/10.1016/j.ifacol.2021.08.353, 19th IFAC Symposium on System Identification SYSID 2021.

[39] Marcus Haywood-Alexander, Nikolaos Dervilis, Keith Worden, Elizabeth J. Cross, Robin S. Mills, Timothy J. Rogers, Structured machine learning tools for modelling characteristics of guided waves, Mech. Syst. Signal Process. 156 (2021) 107628, http://dx.doi.org/10.1016/j.ymssp.2021.107628.

[40] Laura P. Swiler, Mamikon Gulian, Ari L. Frankel, Cosmin Safta, John D. Jakeman, A survey of constrained gaussian process regression: Approaches and implementation challenges, J. Mach. Learn. Model. Comput. 1 (2) (2020) 119–156, http://dx.doi.org/10.1615/JMachLearnModelComput.2020035155.

[41] Jacques H Mclean, Matthew R Jones, Brandon J O'Connell, Eoghan Maguire, Tim J Rogers, Physically meaningful uncertainty quantification in probabilistic wind turbine power curve models as a damage-sensitive feature, Struct. Health Monit. 22 (6) (2023) 3623–3636, http://dx.doi.org/10.1177/14759217231155379.

[42] M.R. Jones, T.J. Rogers, E.J. Cross, Constraining Gaussian processes for physics-informed acoustic emission mapping, Mech. Syst. Signal Process. 188 (2023) 109984, http://dx.doi.org/10.1016/j.ymssp.2022.109984.

[43] Elizabeth J. Cross, Timothy J. Rogers, Thomas J. Gibbons, Grey-box modelling for structural health monitoring: physical constraints on machine learning algorithms, Struct. Health Monit. 2019 (2019) http://dx.doi.org/10.12783/shm2019/32349.

[44] Sören Möller, Clemens Jonscher, Tanja Grießmann, Raimund Rolfes, Investigations towards physics-informed Gaussian process regression for the estimation of modal parameters of a lattice tower under environmental conditions, in: Maria Pina Limongelli, Pier Francesco Giordano, Said Quqa, Carmelo Gentile, Alfredo Cigada (Eds.), Experimental Vibration Analysis for Civil Engineering Structures, Springer Nature Switzerland, Cham, 2023, pp. 401–410, http://dx.doi.org/10.1007/978-3-031-39117-0_41.

[45] Sören Möller, Clemens Jonscher, Tanja Grießmann, Raimund Rolfes, Comparing Gaussian process enhanced grey-box approaches to detect damage in unknown environmental conditions due to climate change, Struct. Health Monit. (2025) http://dx.doi.org/10.1177/14759217241313375.

[46] Eloi Figueiredo, Ionut Moldovan, Pedro Alves, Hugo Rebelo, Laura Souza, Smartphone application for structural health monitoring of bridges, Sensors 22 (21) (2022) http://dx.doi.org/10.3390/s22218483.

[47] Arno Solin, Manon Kok, Know your boundaries: Constraining Gaussian processes by variational harmonic features, in: Kamalika Chaudhuri, Masashi Sugiyama (Eds.), Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics, in: Proceedings of Machine Learning Research, vol. 89, PMLR, 2019, pp. 2193–2202.

[48] Arno Solin, Simo Särkkä, Hilbert space methods for reduced-rank Gaussian process regression, Stat. Comput. 30 (2) (2020) 419–446, http://dx.doi.org/10.1007/s11222-019-09886-w.

[49] Hua-Ping Wan, Wei-Xin Ren, A residual-based Gaussian process model framework for finite element model updating, Comput. Struct. 156 (2015) 149–159, http://dx.doi.org/10.1016/j.compstruc.2015.05.003.

[50] P. Gardner, T.J. Rogers, C. Lord, R.J. Barthorpe, Learning model discrepancy: A Gaussian process and sampling-based approach, Mech. Syst. Signal Process. 152 (2021) 107381, http://dx.doi.org/10.1016/j.ymssp.2020.107381.

[51] Thomas Pinder, Daniel Dodd, GPJax: A Gaussian process framework in JAX, J. Open Source Softw. 7 (75) (2022) 4455, http://dx.doi.org/10.21105/joss.04455.

[52] Benedikt Hofmeister, Marlene Bruns, Raimund Rolfes, Finite element model updating using deterministic optimisation: A global pattern search approach, Eng. Struct. 195 (2019) 373–381, http://dx.doi.org/10.1016/j.engstruct.2019.05.047.

[53] David L. Hunt, Application of an enhanced coordinate modal assurance criterion, in: 10th International Modal Analysis Conference, vol. 1, 1992, pp. 66–71.

[54] Chandula T. Wickramarachchi, Paul Gardner, Jack Poole, Clemens Hübler, Clemens Jonscher, Raimund Rolfes, Damage localisation using disparate damage states via domain adaptation, Data-Centric Eng. 5 (2024) e3, http://dx.doi.org/10.1017/dce.2023.29.

[55] Clemens Jonscher, Benedikt Hofmeister, Tanja Grießmann, Raimund Rolfes, Influence of environmental conditions and damage on closely spaced modes, in: Piervincenzo Rizzo, Alberto Milazzo (Eds.), European Workshop on Structural Health Monitoring, Springer International Publishing, Cham, 2023, pp. 902–911, http://dx.doi.org/10.1007/978-3-031-07322-9_91.

[56] Cristian Guillermo Gebhardt, Benedikt Hofmeister, Christian Hente, Raimund Rolfes, Nonlinear dynamics of slender structures: a new object-oriented framework, Comput. Mech. 63 (2) (2019) 219–252, http://dx.doi.org/10.1007/s00466-018-1592-7.

[57] Daniel James Pitchforth, Matthew Rhys Jones, Samuel John Gibson, Elizabeth Jane Cross, Physically-informed change-point kernels for structural dynamics, 2025, http://dx.doi.org/10.48550/arXiv.2506.11625.

[58] T.J. Rogers, P. Gardner, N. Dervilis, K. Worden, A.E. Maguire, E. Papatheou, E.J. Cross, Probabilistic modelling of wind turbine power curves with application of heteroscedastic Gaussian process regression, Renew. Energy 148 (2020) 1124–1136, http://dx.doi.org/10.1016/j.renene.2019.09.145.

[59] Stefan Wernitz, Benedikt Hofmeister, Clemens Jonscher, Tanja Grießmann, Raimund Rolfes, LUMO - Leibniz university test structure for monitoring, 2021, http://dx.doi.org/10.25835/0027803.