Article:

On the hierarchical Bayesian modelling of frequency response functions

T.A. Dardeno [a,*], K. Worden [a], N. Dervilis [a], R.S. Mills [a], L.A. Bull [b]

[a] *Dynamics Research Group, Department of Mechanical Engineering, University of Sheffield, Sheffield S1 3JD, UK*
[b] *Department of Engineering, University of Cambridge, CB3 0FA, UK*

ABSTRACT

Structural health monitoring (SHM) strategies seek to evaluate, predict, and maintain structural integrity, to improve the safety and design service life of structures in operation. Many of these strategies involve monitoring changes in structural dynamics, as damage can affect modal properties and present as changes in the characteristics of the resonance peaks of the frequency response function (FRF). While recent advances have improved the safety and reliability of structures, a number of challenges remain, impeding the practical implementation and generalisation of these systems. Like damage, benign variations, such as those caused by changes in temperature or other environmental fluctuations, can affect dynamic properties, making it difficult to distinguish between damage and normal operating conditions. In addition, newly-deployed structures can have insufficient data to describe the normal operating conditions (i.e., data *scarcity*), which can impair the development of data-based prediction models. Another common challenge is data loss (i.e., data *sparsity*), which may result from transmission issues, sensor failure, a sample-rate mismatch between sensors, and other causes. Missing data in the time domain will result in decreased resolution in the frequency domain, which can impair dynamic characterisation.

For situations that may benefit from information sharing among datasets, e.g., population-based SHM of similar structures, the hierarchical Bayesian approach provides a useful modelling structure. Hierarchical Bayesian models learn statistical distributions at the population (or parent) and the domain levels simultaneously, to bolster statistical strength among the parameters. As a result, variance is reduced among the parameter estimates, particularly when data are limited. In this paper, a combined probabilistic FRF model is developed for a small population of nominally-identical helicopter blades, using a hierarchical Bayesian structure, to support information transfer in the context of sparse data. The modelling approach is also demonstrated in a traditional SHM context, for a single helicopter blade exposed to varying temperatures, to show how the inclusion of physics-based knowledge can improve generalisation beyond the training data, in the context of scarce data. These models address critical challenges in SHM, by accommodating benign variations that present as differences in the underlying dynamics, while also considering (and utilising), the similarities among the domains.

1. Introduction

The current work is focussed on developing vibration-based structural health monitoring (SHM) strategies, that utilise *data sharing*, domain expertise, and/or physics-based knowledge, to enable information transfer among similar domains. Specifically, this

---

work has involved the development of probabilistic frequency response function (FRF) models, using data collected from healthy, full-scale composite helicopter blades, with consideration for the similarities and differences among the data, and the resulting effects on the characteristics of the resonance peaks.

Benign variations confounding monitoring systems is one of the major challenges in SHM, as even among nominally-identical structures, variations caused by manufacturing differences, ageing parts, and changes in testing conditions can introduce uncertainty in the underlying dynamics. One such example of the effects of manufacturing differences on dynamic properties was presented in [1], where Cawley et al. performed vibration testing on filament-wound carbon fibre-reinforced plastic (CFRP) tubes. Of the 18 tested, six tubes were considered 'normal' (i.e., having the same microstructure, with a ±45° fibre winding angle) [1]. They found that the first and second natural frequencies of the 'normal' tubes varied by as much as 4% [1]. When the remaining 12 tubes were also considered (which had intentional defects in their microstructure, such as slightly misaligned fibres, changes in volume fraction, etc.), the natural frequencies varied by as much as 18% [1]. Likewise, changes in bolt tightness can similarly cause variations in natural frequency. Zhu et al. [2] performed modal testing on an aluminium three-bay space-frame structure with bolted joints, before and after manually loosening several bolts to hand tight. They found that natural frequencies changed by as much as 8%, depending on the mode considered [2].

Global variations, such as changes in ambient temperature, can also affect modal properties. For example, increased temperature may reduce stiffness (depending on the specific material properties of the structure, and the duration/temperature of exposure), which may decrease natural frequency. Colakoglu [3] tested a polyethylene fibre composite beam from −10 °C to 60 °C, and found that the first natural frequency decreased by 12% over the measured temperature range. In addition, for the same polyethylene beam, Colakoglu [3] found that modal damping increased with rising temperature. They concluded that the relationships between temperature and natural frequency, and between temperature and damping, appeared to be functional and monotonic for the polyethylene beam, over the temperature range considered [3]. Similar effects of temperature on stiffness/natural frequency and damping have been noted for a magnesium alloy [4] and a composite honeycomb structure [5]. Accounting for these benign fluctuations is important for the practical implementation and generalisation of SHM technologies, as features commonly used for damage identification may be sensitive to harmless changes as well as damage [6–9]. (Indeed, it is well known that structural damage can reduce stiffness, often manifesting as a reduction in natural frequency. Benign variations can then either mimic or mask damage, depending on whether they exhibit a stiffening or softening effect [7–9]. Similar effects have been noted with damping, and multiple studies have shown that damping tends to increase with damage, particularly with crack growth [10].)

Data *scarcity* and *sparsity* present additional challenges for SHM systems that rely on machine learning. Data scarcity refers to incomplete information regarding the damage- or normal-condition states of structures, particularly those newly in operation, and can impair model training and development. Likewise, sensing networks are prone to data loss (causing sparse data), because of sensor failure caused by harsh environmental conditions or insufficient maintenance. Transmission issues make wireless-sensing networks particularly susceptible to loss, and can be caused by large transmission distances between the sensors and base station [11], software/hardware problems [12], and other issues such as weather changes, interference from nearby devices, or installation difficulties [13]. Furthermore, modern systems that produce large amounts of high-resolution data can suffer losses resulting from a data transfer bottleneck [14]. Significant losses higher than 30% have been reported [12,13,15], and a 0.38% data loss was found to have similar effects on power spectral density (PSD) as 5% additive noise [16]. Differences in sample rate among sensors could have a similar presentation, with the data captured at a lower rate seemingly missing data relative to that captured at a higher rate.

Population-based SHM (PBSHM) addresses these challenges, by supporting knowledge transfer (e.g., damage or normal-condition states) within a population of structures, so that data-rich members can support those with less information. PBSHM strategies differ according to the similarity of the population members. *Heterogeneous* populations [17,18] refer to those comprised of highly disparate members, such as different suspension bridge designs, and may require extensive processing such as domain adaptation prior to information sharing. The current work considers a *homogeneous* population, comprised of nominally-identical members [19], where variabilities in their dynamics result from minor discrepancies in geometry, boundary conditions, and environmental/operational changes. Homogeneous populations can be represented using a general model, called a population *form*, which attempts to capture the 'essential nature' of the population and benign variations among the members [19]. The form was first introduced in [19], where a conventional or single Gaussian process (GP) was applied to frequency response functions (FRFs), to develop a representation for a nominally-identical population of eight degree-of-freedom (DOF) systems. Then, to accommodate greater differences among the nominally-identical members, an overlapping mixture of Gaussian processes (OMGP) [20], was used in [19,21], to infer multivalued wind-turbine power-curve data, with unsupervised categorisation of the data. The OMGP approach [20] was again used in [22] to develop a population form for real and imaginary FRFs, obtained from four nominally-identical, full-scale, composite helicopter blades. Recently, hierarchical modelling was used to improve the predictive capability of simulated [23] and in-service [24] truck-fleet hazard models, and wind-turbine power curves [24]. Specifically, the work presented in [23,24] used *partial pooling*, where data groups are treated as belonging to different domains, and each domain can be considered a realisation from global or population-level distributions. (This approach is in contrast to *complete pooling*, where all data are considered as belonging to a single domain, or *no pooling*, where each domain is treated as fully independent of the other domains). It was shown that when populations of structures were allowed to share correlated information, model uncertainty was reduced [23,24]. In addition, domains with incomplete data were able to borrow statistical strength from data-rich groups [23,24]. In [25], multilevel modelling with partial pooling was used to learn a 2D map of arrival times for waves propagating through a complex plate geometry, via GP regression, using a series of acoustic-emission experiments performed on the same plate but with differing experimental designs. Domain expertise regarding the positive gradient for the expected intertask functions was encoded into the model, via a linear mean function and appropriate priors [25,26]. GP kernel hyperparameters were learnt at the domain level, to allow variation in the response surface among the different tests, and were correlated via a higher-level sampling distribution [25].

1.1. Research aims of the current work

The current work addresses critical SHM challenges, including uncertainties in the underlying dynamics of structures, and data-scarcity/sparsity issues, which can impair generalisation of SHM technologies. In line with the population forms developed in [19,21–24], this paper presents generalised, probabilistic FRF models that were developed for the helicopter blades in [22], using a hierarchical Bayesian structure. Two case studies are presented. The first case used FRFs collected from all four blades, at ambient laboratory temperature, with variations among the blades resulting from manufacturing differences (e.g., small discrepancies in material properties and geometry) and boundary conditions. Limited training data that did not fully characterise the resonance peaks were taken from two of the FRFs, while sufficient training data were taken from the remaining two FRFs, so that information could be shared with the data-poor domains via shared distributions over the parameters. This situation is representative of incomplete data in the time domain, which would reduce the number of spectral lines in the frequency domain. Independent models were generated for comparison, to visualise the variance reduction from the combined model. The second case addressed another challenge, typical in a traditional SHM context, where insufficient information regarding the normal operating condition of a structure can impede model generalisation. In this case, vibration data were collected from a single helicopter blade, at various temperatures, in an environmental chamber. A probabilistic FRF model was again developed using a hierarchical approach with partial pooling, except, for this model, each domain represented a different temperature state. Functional relationships between temperature and natural frequency, and between temperature and damping, were approximated via Taylor series expansion, and polynomial coefficients were learnt at the 'population' level. A subset of temperature-varied FRFs were used to train the model. Higher-level distributions over the modal parameters, and the polynomial coefficients, were then used to generalise to temperature-varied FRFs not used to train the model. Model accuracy was evaluated by comparing the results to FRFs computed via measured vibration data.

1.2. Paper layout

The layout of this paper is as follows. Section 2 summarises existing research related to population-level monitoring of systems. Section 3 outlines the novelty and contribution of the current work. Section 4 presents the theoretical basis for this research, including modal analysis and hierarchical Bayesian modelling techniques. Section 5 briefly describes the datasets used to develop the models. Sections 6 and 7 discuss the models developed and analysis results for the first and second case studies, respectively. Conclusions are presented in Section 8, followed by acknowledgements.

2. Related work

Most literature related to population-level monitoring of engineering structures (or, in a more general SHM context, group-level monitoring of datasets), is focussed on *transfer learning*, which aims to improve predictions in a *target domain* given a *source domain* with more complete information. Some cases have involved fine-tuning the classifiers and weights of a pre-trained convolutional neural network (CNN) according to a new dataset. For example, CNNs have been used to detect cracks, as in [27–30], and other defects (e.g., corrosion, staining), as in [27,31]. Other cases have used domain adaptation (DA), a subcategory of transfer learning, where the source and target domains have the same feature space, and the target domain is mapped onto a shared space. Predictions are then made from a single model. For example, domain-adversarial neural networks have been implemented for condition monitoring of a fleet of power plants [32], and for fault detection in a gearbox and three-story structure [33].

Population-level modelling can also be viewed in the context of multitask learning (MTL), where multiple tasks are solved simultaneously, while considering similarities (and differences) across tasks. In MTL, a combined inference allows sharing of correlated information among domain-specific models, to improve accuracy in data-poor domains [24,34]. In the context of modelling engineering infrastructure, MTL has been used as part of a multioutput Gaussian process (GP) regression to learn correlations between data obtained at adjacent sensors from the Canton Tower, to reconstruct missing information [35]. (Note that using MTL in this context differs from the current work. In [35], missing time-domain data from temperature and acceleration sensors were reconstructed using MTL, and all data were sourced from a single structure. The current work considers frequency-domain data from multiple structures, as in the first case, and incorporates the relationships between temperature and domain-specific tasks into the model structure, as in the second case.) A similar approach was used in [36], where GP regression assisted with missing data recovery from a faulty sensor, by capturing the correlation among the remaining sensors on a hydroelectric dam. Likewise, in [37], GPs were used to transfer information between spatial temperature or pressure profiles at axial stations within an aeroengine.

Hierarchical Bayesian modelling is an MTL framework, where (lower-level) domain-specific tasks are correlated by conditioning on (higher or population-level), *shared* variables. An early introduction of these models to SHM was presented in [38], where a hierarchical Bayesian framework was proposed for accommodating variability in the elastic modulus of simulated bar structures, and in [39], which involved the model updating and uncertainty quantification of a numerical three-story shear-building model in the presence of environmental variations. Likewise, in [40,41], a multilevel hierarchical Bayesian model was developed to identify damage on a single structure by inferring stiffness loss from changes in modal parameters, given noisy, incomplete modal data. Modal parameters were learnt for different damage scenarios using a series of coupled linear regressions to approximate the eigenvalue problem [40,41]. In [41], the similarities between acceleration responses at adjacent sensors were exploited for data-loss recovery, by modelling the measured signal as a linear combination of basis functions. Both cases presented in [40,41] considered data from a single structure. (Note that the approaches presented in [38–41] differ from the current work. The work presented in the current paper, in the first case study, models the entire (band-limited) FRF from *multiple* structures, using a likelihood function based on

modal parameters. The second case models the entire FRF using measurements at various temperatures from a single structure, and incorporates functional relationships between temperature and the modal parameters, enabling predictions at 'unseen' temperature conditions.) Recently, [42] used hierarchical Bayesian modelling to identify the electromechanical properties of energy harvesters, via individual and combined inference of experimentally-obtained FRFs. Population-level parameters were compared for partial- and complete-pooling approaches, successfully demonstrating that partial pooling results in improved estimates of population-level distributions compared to a classical Bayesian approach. The current work also utilises a hierarchical approach to probabilistic FRF modelling, but instead, the focus is on *knowledge transfer* - whereby one or more members of a population or group have markedly fewer data. The current work also increases the complexity of the population (parameter) models, to move beyond exchangeable tasks, and predict the model parameters for previously-unobserved experiments. (In other words, the current work allows for the prediction of unseen models, as well as data.)

Hierarchical Bayesian models were again used in [43], to estimate corrosion rates via data from multiple sensors on the same test structure, to support decision-making in the absence of complete data. In [44] hierarchical Bayesian models were developed using strain measurements from multiple tensile tests to inform the material properties of the samples, in a manner that considered the inherent variability in material properties among the samples and uncertainties related to experimental repeatability. Likewise, in [45], hierarchical Gaussian mixture models were used for combined inference in a simulated population of structures, for the purpose of damage identification. Another recent development was presented in [46], where a physics-informed GP model was developed using measured time-domain vibration data, with accommodation for temporal variability accomplished by partitioning the data into multiple subsegments, and applying a hierarchical Bayesian approach.

3. Novelty and contribution

Unlike previous efforts, the current work is concerned with developing a probabilistic FRF model, using an FRF equation based on modal parameters (natural frequency, mode shape, and damping) as the likelihood function. As with [23–25], a hierarchical Bayesian (or MTL) framework was used. However, the focus is data sharing in the presence of low-resolution frequency information (i.e., given *sparse* data). The inability to localise both time and frequency to a fine resolution, per the uncertainty principle, means that missing time-domain data (e.g., acceleration, velocity) measured from SHM sensors will result in fewer spectral lines in the frequency domain. This decreased frequency resolution could impair proper characterisation/identification of modal peaks, which may be features of interest in a damage-detection application. By using a combined-inference approach, structures whose FRFs have many spectral lines in a band of interest can lend statistical support to those whose FRFs are limited by missing data. In addition, the current work incorporated functional relationships to describe changes in environmental conditions. This inclusion of functional relationships allows for extrapolation to temperature states not used in model training, which increases the amount of normal-condition information available, to address data-scarcity challenges. In contrast to [25], parameters were learnt over the experimental campaign, rather than hyperparameters, giving greater physical interpretability of the results. In contrast to previous work, the current work focusses on information transfer for modal analysis and meta-analysis prediction for future models, rather than data pooling alone. Beyond conventional multilevel modelling and partial pooling, the intent is to learn how parameters vary over the population, even in the absence of complete data - an area that has not yet been addressed in modal analysis. Again, the current work allows for the prediction of new models, in addition to unseen data.

4. Background theory

In this work, combined probabilistic FRF models were developed using a hierarchical Bayesian modelling approach, to support SHM research. This section provides an overview of the methodologies used, including hierarchical multilevel modelling from a Bayesian perspective, linear modal analysis, FRF estimation, and model evaluation.

4.1. Hierarchical Bayesian modelling

Hierarchical models can be used to make combined inferences, whereby domains are treated as separate; but, at the same time, it is assumed that each domain is a realisation from a common latent model. This modelling structure involves *partial pooling*, and is beneficial in that population-level distributions are informed by the full dataset, comprised of multiple domains. In partial pooling, certain parameters are permitted to vary between domains (i.e., *varying parameters*); which are correlated by conditioning on parent variables at the population level. In the current work, the natural frequency would be a varying parameter. Other parameters can be considered *shared* among members of a population (e.g., additive noise) and are learnt at the population level (these shared variables can still be sampled from parent distributions, which are also learnt at the population level). In contrast, a complete-pooling approach would consider all population data as having originated from a single source, while a no-pooling approach would involve fitting a single domain independently from the other domains.

Hierarchical models with partial pooling can be applied in a variety of SHM contexts, where a system can benefit from information sharing while preserving the uniqueness of the individual data sources. However, these models are particularly useful for PBSHM, and link intuitively to the concept of the population *form*, which encompasses both the essential nature and variabilities among the members of a homogeneous population. Because parameters are allowed to vary at the domain level (as opposed to complete pooling), this approach can represent benign variations within a population. In addition, population-level variables are informed by the full dataset, rather than data from a single domain. This increase in statistical power is especially important in situations where
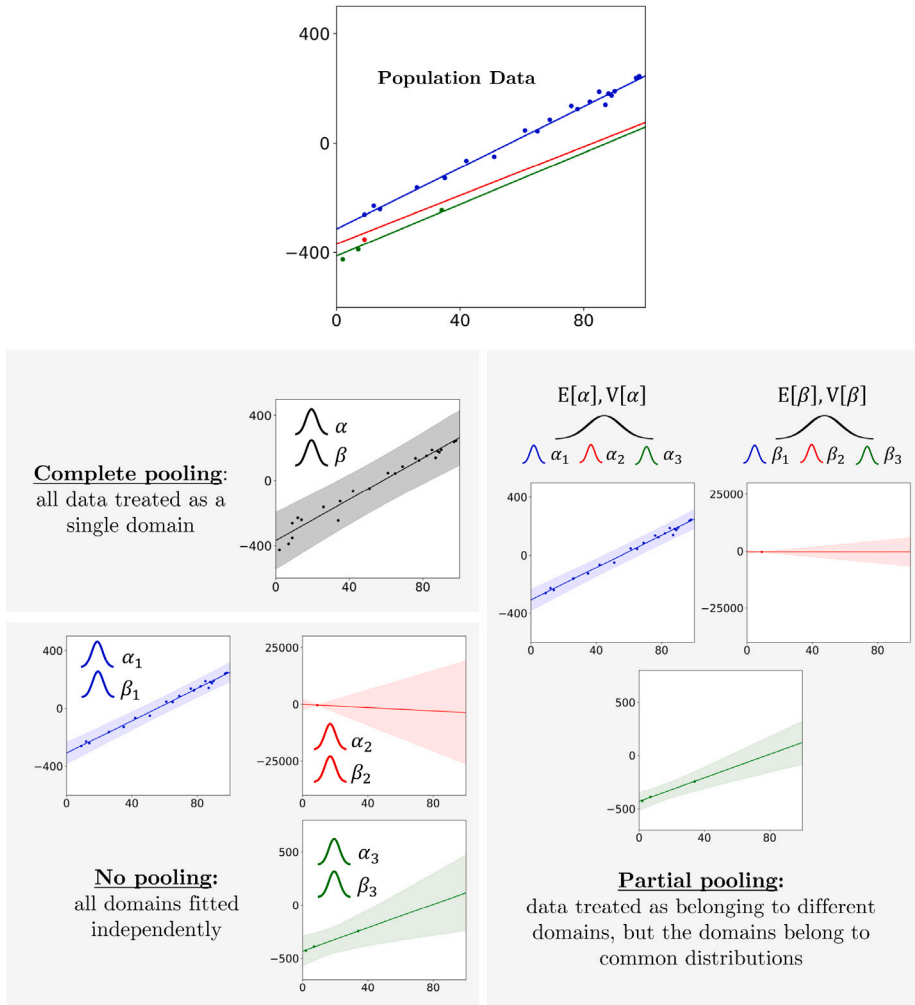
Fig. 1. Comparison of data-pooling approaches, using a simple linear regression example.

one or more domains have limited data [23,24,47]. In such cases, parameters from the data-poor domains exhibit shrinkage towards the population mean (therefore borrowing information from the other domains), which tightens the parameter variance [23,24,47]. The differences among the data-pooling techniques are shown graphically in Fig. 1, using a simple linear regression example.

From a general perspective, data from a population comprised of K groups can be denoted via,

$$\left\{ \mathbf{x}_k, \mathbf{y}_k \right\}_{k=1}^{K} = \left\{ \left\{ x_{ik}, y_{ik} \right\}_{i=1}^{N_k} \right\}_{k=1}^{K} \tag{1}$$

where $\mathbf{y}_k$ is the target response vector for inputs $\mathbf{x}_k$, and $\left\{ x_{i,k}, x_{i,k} \right\}$ are the ith pair of observations in group k [24]. Each group is comprised of N_k observations, giving a total of $\sum_{k=1}^{K} N_k$ observations [24]. The objective is to learn a set of K predictors (one for each domain), related to the regression task, where the tasks satisfy,

$$\left\{ y_{ik} = f_k \left(x_{ik} \right) + \epsilon_{ik} \right\}_{k=1}^{K} \tag{2}$$

In other words, for each ith observation, the output is determined by evaluating one of K latent functions, $f_k \left(x_{i,k} \right)$, plus additive noise, $\epsilon_{i,k}$ [24].

While each of the k groups can be learned independently, a combined inference can be used to take advantage of the full $\sum_{k=1}^{K} N_k$ population dataset. For example, consider a population that can be expressed using K linear regression models, as in Fig. 1,

$$\left\{ \mathbf{y}_k = \boldsymbol{\alpha}_k + \boldsymbol{\beta}_k \mathbf{x}_k + \boldsymbol{\epsilon}_k \right\}_{k=1}^{K} \tag{3}$$

where $\boldsymbol{\alpha}_k$ and $\boldsymbol{\beta}_k$ are the intercept and slope for domain k, respectively; $\mathbf{x}_k$ is a vector of inputs for the kth domain, with length N_k; and $\boldsymbol{\epsilon}_k$ is the noise vector for the kth domain, with length N_k, and is assumed to be normally-distributed. Then, the likelihood of
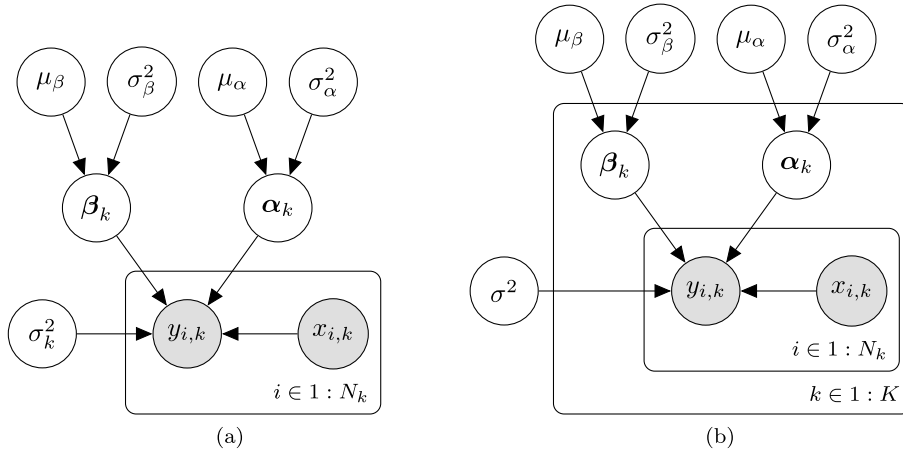
Fig. 2. DGM of linear regression model, (a) independent (no pooling) and (b) partial pooling.

the target response vector is given as,

$$\mathbf{y}_k | \mathbf{x}_k \sim \mathcal{N} \left(\boldsymbol{\alpha}_k + \boldsymbol{\beta}_k \mathbf{x}_k, \sigma_k^2 \right) \tag{4}$$

A shared hierarchy of prior distributions can be placed over the slopes and intercepts for the groups $k \in \{1, \dots, K\}$, in line with a Bayesian framework. To allow information to flow between groups, parent nodes $\{\boldsymbol{\mu}_\alpha, \sigma_\alpha^2\}$ and $\{\boldsymbol{\mu}_\beta, \sigma_\beta^2\}$ can be learned at the population level. Note that sometimes, it may be appropriate to learn certain parameters at the population level rather than the domain level. If, for example, the same hardware was used for data collection within the population, one could assume a global noise variance (σ^2) that is the same for each domain. Consider the directed graphical models (DGMs) shown in Figs. 2(a) and 2(b). Fig. 2(a) shows the DGM for the linear regression example, given an independent (no-pooling) approach. Each model is learnt independently and no information is shared. Slopes, intercepts, and noise variance are learnt individually for each k domain, but there is no plate surrounding them to denote repetition over the K domains. While the slopes and intercepts are sampled from shared parent nodes $\{\boldsymbol{\mu}_\alpha, \sigma_\alpha^2\}$ and $\{\boldsymbol{\mu}_\beta, \sigma_\beta^2\}$, these nodes are informed only by the data from a *single* domain and any priors. In contrast, Fig. 2(b) shows the DGM given a partial-pooling approach. The slopes and intercepts are indexed by k, and plate notation is used to show that these nodes are repeated. Once again, shared parent nodes are outside of the plates and not indexed by k, meaning that these are population-level variables, and information is permitted to flow between these and the domain-specific parameters. In other words, the parent nodes are now informed by the data from *multiple* domains and prior information. The noise variance (σ^2) is shared among the domains, at the population level, as indicated by the absence of a k subscript, and its exclusion from the K plate.

The current work considers a hierarchical probabilistic FRF model, that uses an accelerance FRF estimate (based on modal parameters), as the mean of the likelihood function. A brief introduction to modal analysis and FRF estimation is provided.

4.2. Modal analysis and FRF estimation

The equation of motion for a multiple DOF system can be written as,

$$\mathbf{M}\ddot{\mathbf{u}}(t) + \mathbf{C}\dot{\mathbf{u}}(t) + \mathbf{K}\mathbf{u}(t) = \mathbf{z}(t) \tag{5}$$

where $\ddot{\mathbf{u}}(t)$, $\dot{\mathbf{u}}(t)$, $\mathbf{u}(t)$, and $\mathbf{z}(t)$ are acceleration, velocity, displacement, and force, respectively. In most cases, the mass $\mathbf{M}$, damping $\mathbf{C}$, and stiffness $\mathbf{K}$ matrices are coupled. For linear systems, and in the absence of viscous damping, the equation of motion can be decoupled, such that the system is represented by multiple single degree-of-freedom (SDOF) oscillators. This decoupling is performed via the eigenvalue expression,

$$[\mathbf{K} - \omega_n^2 \mathbf{M}]\boldsymbol{\Psi} = 0 \tag{6}$$

and yields the natural frequencies in radians, ω_n, and mode shapes, $\boldsymbol{\Psi}$, of the system.

The physical equation of motion can then be cast in modal space to give the uncoupled modal equations, with modal coordinates $\mathbf{p}(t)$, written as,

$$\boldsymbol{\Psi}^{\mathrm{T}}\mathbf{M}\boldsymbol{\Psi}\ddot{\mathbf{p}}(t) + \boldsymbol{\Psi}^{\mathrm{T}}\mathbf{C}\boldsymbol{\Psi}\dot{\mathbf{p}}(t) + \boldsymbol{\Psi}^{\mathrm{T}}\mathbf{K}\boldsymbol{\Psi}\mathbf{p}(t) = \boldsymbol{\Psi}^{\mathrm{T}}\mathbf{z}(t) \tag{7}$$

where,

$$\mathbf{u}(t) = \boldsymbol{\Psi}\mathbf{p}(t) \tag{8}$$

FRFs aid visualisation of the natural frequency components of a system, and are computed by normalising the response signal at a given location to the excitation force. This work used the H_1 estimator to compute FRFs. For a response at location h resulting from excitation at j, the H_1 estimator is computed as,

$$\mathbf{H}_{hj}(\omega) = \frac{\mathbf{G}_{zu}(\omega)}{\mathbf{G}_{zz}(\omega)} \tag{9}$$

where,

$$\mathbf{G}_{zu}(\omega) \overset{\Delta}{=} \mathbb{E}[\mathbf{S}_{u_h}(\omega)\mathbf{S}^*_{z_j}(\omega)]$$
$$\mathbf{G}_{zz}(\omega) \overset{\Delta}{=} \mathbb{E}[\mathbf{S}_{z_j}(\omega)\mathbf{S}^*_{z_j}(\omega)] \tag{10}$$

and,

$$\mathbf{S}_{z_j}(\omega) \overset{\Delta}{=} \mathcal{F}[\mathbf{z}_j(t)]$$
$$\mathbf{S}_{u_h}(\omega) \overset{\Delta}{=} \mathcal{F}[\mathbf{u}_h(t)] \tag{11}$$

The asterisk $*$ denotes complex conjugation, ω is frequency in radians, and $\mathcal{F}$ is a discrete Fourier transform (this work used a fast Fourier transform or FFT). The input force in the time domain at location j is $\mathbf{z}_j(t)$, and the output response (i.e., acceleration, velocity, or displacement) in the time domain at h is $\mathbf{u}_h(t)$.

With assumed linear behaviour and proportional damping, the (complex) accelerance FRF (i.e., given acceleration response data) can also be estimated using modal parameters,

$$\mathbf{H}_{hj}(\boldsymbol{\omega}) = -\boldsymbol{\omega}^2 \sum_{m=1}^{M} \frac{A^{(m)}_{hj}}{\omega^2_{nat,m} - \boldsymbol{\omega}^2 + 2i\zeta_m \boldsymbol{\omega}\omega_{nat,m}} \tag{12}$$

where $A^{(m)}_{hj}$ is the residue for mode m, defined as the product of the mass-normalised mode shapes at locations h and j ($A^{(m)}_{hj} = \psi_{hm}\psi_{jm}$) [48]. The natural frequency associated with mode m is $\omega_{nat,m}$, and the modal damping associated with mode m is ζ_m [48]. The real and imaginary parts of the FRF can be computed independently, by multiplying Eq. (12) by its complex conjugate. This work considers FRFs from multiple domains (different structures in Case 1, and a single structure in Case 2 with data obtained at varying temperatures), with each domain indexed by k. In each case, only one FRF is assigned to each domain, from a given measurement location (so, subscripts h and j can be neglected). Thus, the real and imaginary components of the FRF from the kth domain, given a vector of frequency inputs $\boldsymbol{\omega}_k$, can be estimated via,

$$\text{real}\left[\mathbf{H}_k\left(\boldsymbol{\omega}_k\right)\right] = -\boldsymbol{\omega}^2_k \sum_{m=1}^{M} \frac{A_m \left(\omega^{2(k,m)}_{nat} - \boldsymbol{\omega}^2_k\right)}{\omega^{4(k,m)}_{nat} + \boldsymbol{\omega}^4_k + 2\boldsymbol{\omega}^2_k \omega^{2(k,m)}_{nat}\left(2\zeta^2_{(k,m)} - 1\right)} \tag{13}$$

$$\text{imag}\left[\mathbf{H}_k\left(\boldsymbol{\omega}_k\right)\right] = -\boldsymbol{\omega}^2_k \sum_{m=1}^{M} \frac{-2A_m i\zeta_{(k,m)}\boldsymbol{\omega}_k \omega^{(k,m)}_{nat}}{\omega^{4(k,m)}_{nat} + \boldsymbol{\omega}^4_k + 2\boldsymbol{\omega}^2_k \omega^{2(k,m)}_{nat}\left(2\zeta^2_{(k,m)} - 1\right)} \tag{14}$$

Note that residues, A_m, are not indexed by k in Eq. (13). For the models presented in the case studies below, mode shapes were shared among the domains, to address identifiability concerns during sampling, and because mode shapes are often less sensitive to global variations compared to other modal parameters.

The current work has focussed on the real part of the FRF, and Eq. (13) provided the mean of the likelihood function for the models developed. The residues, modal damping, and natural frequencies for each domain were learnt as parameters for the hierarchical model, with distributions over these parameters learnt as hyperparameters. Learning the real part of the FRF was sufficient for demonstrating the proposed technology. However, the imaginary part could be learnt using the same methods, with Eq. (14) providing the likelihood function, or could be inferred (at least in part), by exploiting the causal relationship between the real and imaginary components of the FRF [48].

4.3. Model evaluation for generalisation beyond training data (Case 2)

The normalised mean-squared error (NSME) was calculated to evaluate the accuracy of the extrapolation to temperatures beyond the training data, for the second case presented herein, via

$$\text{NMSE} = \frac{100}{M\sigma^2_{\mathbf{y}}} \sum_{i=1}^{M} \left(y_i - y^*_i\right) \tag{15}$$

where M is the number of test data points, y is the test data, $\sigma^2_{\mathbf{y}}$ is the variance of the test data, and y^* is the predicted function. Normalising to the test data variance allows for comparison of results on a consistent scale, regardless of signal magnitude. According to convention, an NMSE less than 5% suggests that the model fits the data well.

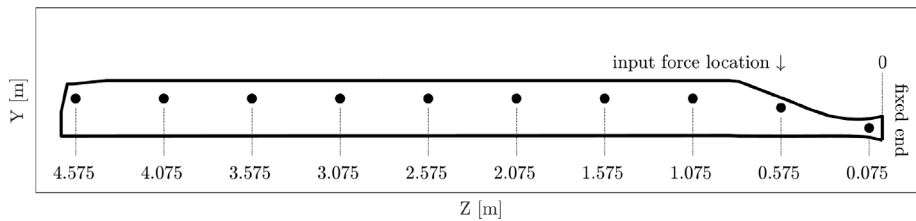**Fig. 3.** Helicopter blade in a substantiated wall mount.

Fig. 4. Sensor locations on the helicopter blades.

5. Dataset summary

The population dataset was comprised of vibration data collected from four healthy, nominally-identical, full-scale composite[1] blades from a Gazelle helicopter (referenced in this paper as Blades 1–4). Data were collected using Siemens PLM LMS SCADAS hardware and software at the Laboratory for Verification and Validation[2] (LVV) in Sheffield, UK. The first case used FRFs calculated from data that were collected on all four blades at ambient laboratory temperature, as described in [22]. The second case used FRFs calculated from data that were collected on a single blade (Blade 1) at multiple temperatures in an environmental chamber.
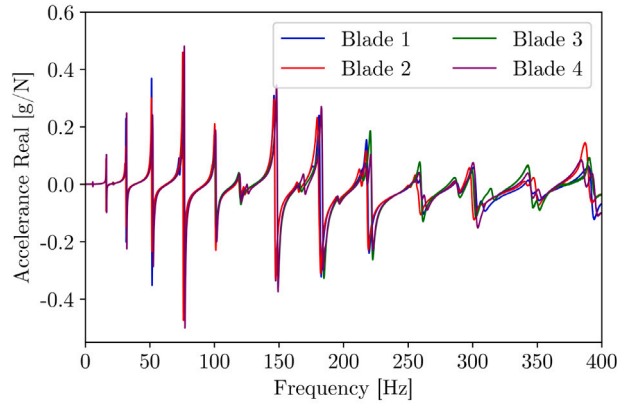
5.1. Data collection at ambient laboratory temperature

The first case used data collected at ambient laboratory temperature, with the blades in a fixed-free boundary condition, which was approximated by placing the root end of each blade in a substantiated strong-wall mount. Ten uniaxial 100 mV/g accelerometers were placed along the length of the underside of each blade. Note that the same accelerometers were used on each blade, and care was taken to ensure that they were attached to approximately the same locations on each blade. An electrodynamic shaker with force gauge was mounted to a fixture bolted to the laboratory floor and attached to the blade 0.575 metres from the root. The shaker was attached to the underside of the blade in the flapwise direction. A continuous random excitation was generated in LMS (note: LMS refers to Siemens PLM LMS SCADAS hardware and software) and applied to excite the blade up to 400 Hz, with a step size of 4.88e−02 Hz. Throughput time data were collected for each test, and the data were divided into 20 blocks. Hanning windows were applied and FRFs were computed for each time block, which were then averaged in the frequency domain. The experimental setup is shown in Fig. 3, and the accelerometer positions on the blades are shown in Fig. 4.

To demonstrate the proposed technique and reduce the number of model hyperparameters (thus minimising model complexity and reducing computation time), a narrow frequency band was selected around two lightly-damped, well-separated modes. Specifically, the chosen bandwidth was between 24 and 61 Hz, with the fourth and fifth bending modes of the blades dominating the response in this band. Although other modes appear to have a small influence in this band, a 2DOF assumption was imposed. (This assumption results in smoothing of the FRF over the band, and might result in some loss of interpretability, but is acceptable for these preliminary analyses). The real part was modelled as a probabilistic FRF, using the FRF estimate from Eq. (13) as the mean of the likelihood function, as described in Case 1, presented in Section 6 of this paper. The real parts of the averaged FRFs for each blade, at the second accelerometer from the blade root (corresponding to the drive-point location), are shown in Figs. 5(a) and 5(b). Fig. 5(a) shows the full measured bandwidth, and Fig. 5(b) shows the FRF in the bandwidth of interest, between 24 and 61 Hz.

---

[1] Although the exact internal geometry of the blades is unknown, specifications for Gazelle helicopter blades have indicated that they are comprised of steel, fibreglass, and a honeycomb core [49].

[2] https://lvv.ac.uk/.
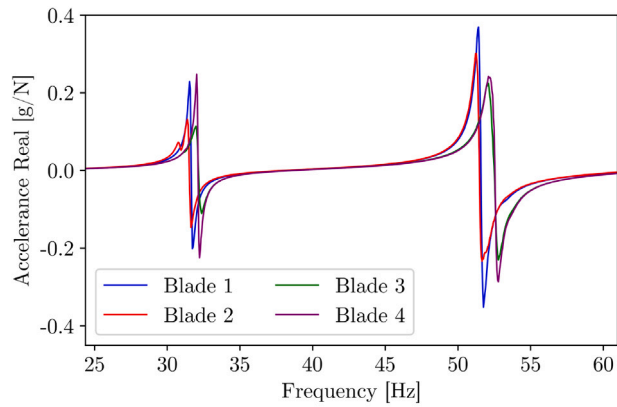
Fig. 5. Real components of the helicopter blades FRFs, (a) full bandwidth and (b) bandwidth of interest, at the drive-point location, from vibration data collected at ambient laboratory temperature.

Figs. 5(a) and 5(b) show increasing variability with respect to frequency, which is an expected result, as higher-frequency modes are more sensitive to small physical changes than lower-frequency modes. For modes less than 80 Hz, the maximum frequency difference among the blades was approximately 2.5 Hz; for modes greater than 80 Hz, the maximum frequency difference was approximately 6.3 Hz. Note the grouping visible at several of the peaks, where Blades 1 and 2 appear closely aligned in frequency while Blades 3 and 4 appear closely aligned. These results are quite relevant for PBSHM. All of the helicopter blades are healthy, and represent a normal-condition state of the population. Consider a situation where only FRFs from one of the groupings are available for training a model (or FRFs from the other groups are missing data). The normal condition could be heavily biased towards the training set, and incoming FRFs could be flagged as damaged, even if they are healthy. Further details regarding the data collection and processing for these tests can be found in [22].

5.2. Data collection at various temperatures in an environmental chamber

The second case used data from Blade 1, collected at temperatures ranging from −20 to 30 °C in increments of 5 °C in an environmental chamber. The blade was tested in an approximately fixed-free boundary condition, with the blade root mounted on a fixture that was substantiated with a large concrete block. These tests used the same accelerometers, sensor layout, data-acquisition parameters, and processing methods as the previous tests at ambient laboratory temperature. The shaker was mounted to the bottom plate of the test fixture, and the shaker and force gauge were connected to the underside of the blades to excite the blades in the flapwise direction. A thermal jacket was used to protect the shaker when testing at lower temperatures. Prior to data collection, the environmental chamber was set to each of the temperatures of interest, after which the blades were allowed to soak for at least two hours to reach the desired temperature. Throughput force and acceleration data were collected for each test. The data were then divided into 20 blocks, and a Hanning window was applied to each data block. FRFs were then computed for each block, and averaged in the frequency domain. The experimental setup, with the helicopter blade inside the environmental chamber, is shown in Fig. 6.
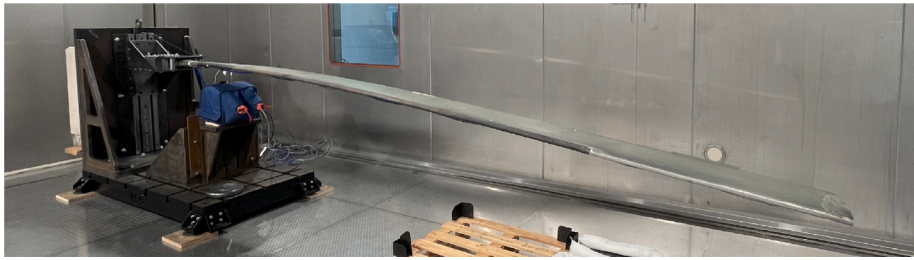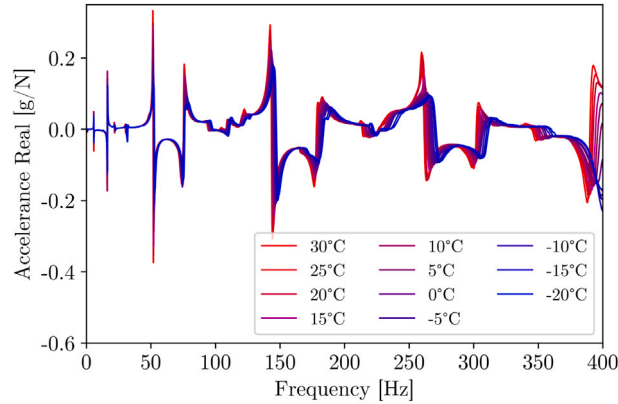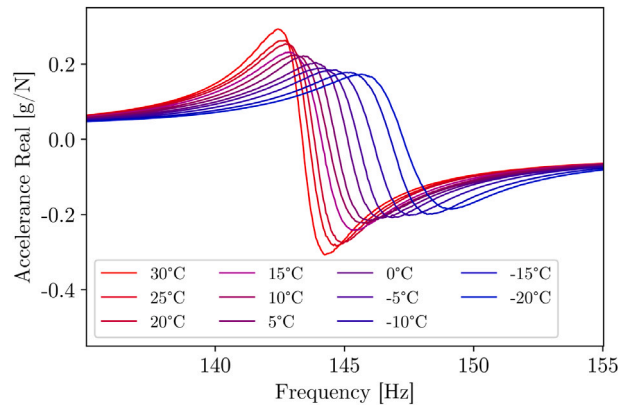
Fig. 6. Helicopter blade in an environmental chamber.

(a)

(b)

Fig. 7. Real components of the helicopter blades FRFs, (a) full bandwidth and (b) bandwidth of interest, at the fourth accelerometer from the blade tip, from vibration data collected at various temperatures in an environmental chamber.

To minimise the number of hyperparameters in the model and reduce computation time, a narrow frequency band was selected around a lightly-damped, well-separated modal peak. Specifically, the chosen bandwidth was between 135 and 155 Hz, with a higher-order bending mode dominating the response in this band. An SDOF assumption was imposed. (Again, this assumption was considered acceptable for these preliminary analyses, to avoid data-separability issues. Likewise, the coefficients of the functions that relate temperature fluctuations to changes in modal parameters will likely be mode-dependent.) FRFs captured at temperatures -10, -5, 10, and 25 °C provided the training data. The remaining temperature-varied FRFs were used to evaluate the extrapolation results to other temperatures. As with the previous case, the real part was modelled as a probabilistic FRF, using the FRF estimate from Eq. (13) as the mean of the likelihood function. The real parts of the averaged FRFs for each temperature, from the fourth accelerometer from the blade tip, are shown in Figs. 7(a) and 7(b). Fig. 7(a) shows the full measured bandwidth, and Fig. 7(b) shows the FRF in the bandwidth of interest, between 135 and 155 Hz.

Figs. 7(a) and 7(b) show a proportional decrease in frequency corresponding to each incremental temperature increase, with discrepancies more noticeable at the higher modes than the lower modes. These results are expected, as the higher-frequency modes are more sensitive to environmental and other changes. The maximum frequency difference among the tests was approximately 15.3 Hz (for modes less than 400 Hz), found in the band between 335 and 375 Hz, as obtained via peak-picking. In the same band, the average frequency difference for each 5 °C increment was approximately 1.5 Hz. For the mode at approximately 145 Hz shown in Fig. 7(b), the maximum frequency difference was approximately 3.8 Hz, and the average frequency difference for each 5 °C increment was approximately 0.38 Hz. Further details regarding the data collection and processing for these tests can be found in [50].

6. Case 1: population-based modelling of FRFs for nominally-identical structures

The first case used FRF data from four nominally-identical helicopter blades, collected at ambient laboratory temperature. A population *form* for the FRFs of the helicopter blades was developed using the aforementioned hierarchical (partial-pooling) approach. Models were developed using the probabilistic programming language Stan. Analyses were performed using MCMC, via the no U-turn (NUTS) implementation of Hamiltonian Monte Carlo (HMC) [51,52]. HMC uses approaches based in differential geometry to generate transitions that span the full marginal variance, which allows the algorithm to accommodate large curvature changes in the target distribution (which are common with hierarchical models) [52].

6.1. Model development

The population of helicopter blade FRFs, with frequency inputs, $\boldsymbol{\omega}_k$, and accelerance outputs, $\mathbf{H}_k$, was re-written from the general representation in Eq. (1),

$$\{\boldsymbol{\omega}_k, \mathbf{H}_k\}_{k=1}^K = \left\{ \{\omega_{ik}, H_{ik}\}_{i=1}^{N_k} \right\}_{k=1}^K \tag{16}$$

where $\{\omega_{ik}, H_{ik}\}$ was the ith pair of observations in domain k. Then, considering only the real component of the FRF, Eq. (2) was re-written as,

$$\{\text{real}\left[H_{ik}\right] = f_k\left(\omega_{ik}\right) + \epsilon_{ik}\}_{k=1}^K \tag{17}$$

where for each ith observation, the output was determined by evaluating one of K latent functions, $f_k\left(\omega_{i,k}\right)$, plus additive noise, $\epsilon_{i,k}$. For this work, there were four helicopter blades and four domains. Therefore, the population model included four latent functions (i.e., $K = 4$). To accommodate outliers in the FRF data, the (real) FRFs were modelled probabilistically with an assumed Student's t-distributed likelihood,

$$\text{real}\left[\mathbf{H}_k\left(\boldsymbol{\omega}_k\right)\right] \sim \mathcal{T}\left(\nu, \mathbf{f}_k\left(\boldsymbol{\omega}_k\right), \sigma^2\right) \tag{18}$$

where $\mathbf{f}_k\left(\boldsymbol{\omega}_k\right)$ was equal to the real component of the FRF, calculated using modal parameters via Eq. (13). For both the independent and partial-pooling models, the degrees-of-freedom parameter, ν, was assumed to be four (4). Setting ν to a constant was necessary to stabilise the independent (no-pooling) models for the data-poor domains, to avoid the sampler falsely identifying the modal peaks as noise. (Note that struggling to fit the ν parameter was not a problem for the data-rich independent models or for the partial-pooling model, only for the data-poor independent models.) The additive noise variance, σ^2, of the FRFs was assumed to be the same for each blade. This assumption was reasonable, as the same data acquisition system and sensors were used among the different tests. Note that $\boldsymbol{\omega}_k$ was permitted to vary depending on the given FRF. This allowed for flexibility, to consider FRFs with different numbers of spectral lines or missing data points, and to represent population uncertainty.

Natural frequencies and modal damping were learnt at the domain level, and allowed to vary among the different helicopter blades. Residues were shared among the different domains and learnt at the population level, to mitigate model identifiability issues. Shared, population-level prior distributions (with hyperpriors) were also placed over the modal parameters to capture/infer the similarity among the FRFs. Domain-level natural frequencies, $\boldsymbol{\omega}_{nat} = \{\{\omega_{nat}^{k,m}\}_{m=1}^2\}_{k=1}^4$, were sampled from a truncated normal parent distribution, with higher-level expectation and variance sampled from truncated normal distributions,

$$\begin{aligned} \boldsymbol{\omega}_{nat} &\sim \mathcal{TN}\left(\boldsymbol{\mu}_{\omega_{nat}}, \boldsymbol{\sigma}^2_{\omega_{nat}}\right) \\ \boldsymbol{\mu}_{\omega_{nat}} &\sim \mathcal{TN}\left([190, 335], [5^2, 5^2]\right) \\ \boldsymbol{\sigma}^2_{\omega_{nat}} &\sim \mathcal{TN}\left([5, 5], [5^2, 5^2]\right) \end{aligned} \tag{19}$$

Note that hyperpriors for $\boldsymbol{\omega}_{nat}$ are shown in rad/s. Domain-level damping, $\boldsymbol{\zeta} = \{\{\zeta_{k,m}\}_{m=1}^2\}_{k=1}^4$, was sampled from a beta parent distribution, with higher-level shape parameters sampled from truncated normal distributions,

$$\begin{aligned} \boldsymbol{\zeta} &\sim \mathcal{B}\left(\boldsymbol{\alpha}_\zeta, \boldsymbol{\beta}_\zeta\right) \\ \boldsymbol{\alpha}_\zeta &\sim \mathcal{TN}\left([6, 6], [0.5^2, 0.5^2]\right) \\ \boldsymbol{\beta}_\zeta &\sim \mathcal{TN}\left([500, 500], [50^2, 50^2]\right) \end{aligned} \tag{20}$$
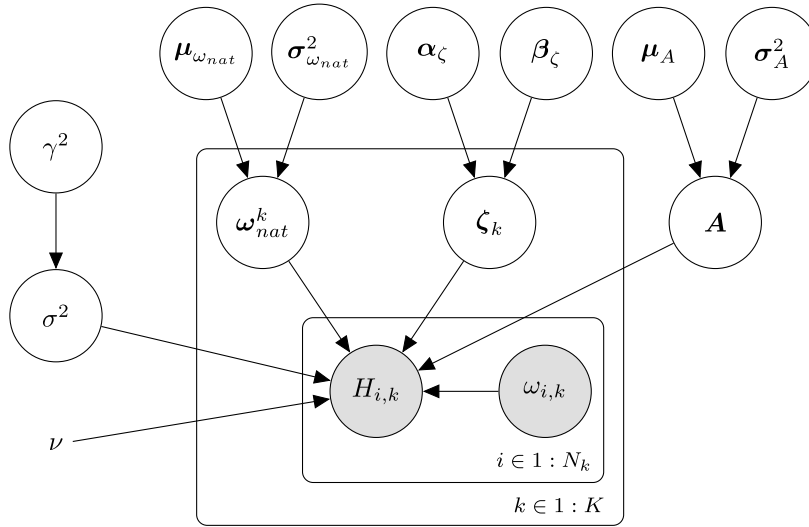
Fig. 8. Graphical representation of the hierarchical FRF model developed for the first case.

Likewise, shared modal residues, $\boldsymbol{A} = \{A_m\}_{m=1}^2$, were sampled from a normal parent distribution, with higher-level expectation and variance sampled from normal and truncated normal distributions, respectively,

$$\boldsymbol{A} \sim \mathcal{TN}\left(\boldsymbol{\mu}_A, \boldsymbol{\sigma}_A^2\right)$$
$$\boldsymbol{\mu}_A \sim \mathcal{N}\left([\text{-0.004, -0.004}], [0.0035^2, 0.0035^2]\right) \tag{21}$$
$$\boldsymbol{\sigma}_A^2 \sim \mathcal{TN}\left([0.003, 0.003], [0.003^2, 0.003^2]\right)$$

For simplicity, normal distributions were chosen for most of the parameters (or truncated normal distributions, if the parameter was restricted to be positive), although other distributions could be used. A beta distribution was chosen for damping, as beta distributions have support $x \in [0, 1]$, which is suitable for lightly-damped systems. Hyperpriors were chosen based on the physics that could be interpreted by visual inspection of the training data (e.g., natural frequency), or by fitting the data-rich domains independently (discussed below).

Shared noise variance, σ^2, was sampled from a half-Cauchy parent distribution, with higher-level scale parameter, γ^2, sampled from a truncated normal distribution,

$$\sigma^2 \sim \mathcal{C}\left(0, \gamma^2\right)$$
$$\gamma^2 \sim \mathcal{TN}\left(0.01, 0.05^2\right) \tag{22}$$

The half-Cauchy was considered an appropriate prior distribution for the shared noise term, as it peaks at zero, but has a long right tail, to allow for some flexibility in the assumptions [53]. (Note that, in practice, measurement noise would be expected to be fairly small given that FRFs are typically averaged.) A graphical model displaying the parameter hierarchy is shown in Fig. 8.

The models were trained using data with added noise, to avoid over-fitting small deviations/lightly-participating modes within the band of interest. This was accomplished by adding zero-mean Gaussian noise, with variance equal to 10% of the absolute maximum amplitude of the data from Blade 1, to the 20 time-data blocks prior to computing FRFs. Then, the time data were subsampled such that there were two data-rich domains (Blades 1 and 2), and two data-poor domains (Blades 3 and 4). (Note that the data-rich groups were subsampled as well. This procedure was adopted to reduce the dimensionality of the training data, although there were enough training points to characterise each mode within the band of interest for these blades.) FRFs were then computed and averaged, and truncated to the desired bandwidth. This procedure resulted in 281, 207, 11, and 31 training points for Blades 1, 2, 3, and 4, respectively.

The intent was that the two data-rich FRFs would lend statistical support to the sparser domains, thereby reducing uncertainty compared to an approach without pooling. (Note that partial pooling is also applicable to other situations, such as having only a single data-rich member, as briefly explored in Appendix.) Recall the grouping of the peaks from the FRFs shown in Figs. 5(a) and 5(b). Both data-rich FRFs belonged to one of the groupings; likewise, the two data-poor FRFs belonged to the other grouping. This represents a challenging situation where the model can be heavily biased towards the available training data. The hierarchical/partial-pooling modelling structure allows the target distributions to be informed by the data (for very data-poor domains, this will of course be limited), but with the help of the full population dataset (including the data-rich domains). In addition, the subsampling and added noise resulted in some outliers in the data, creating a challenging inference problem.

The Stan HMC sampler was run using four chains, with a target average proposal acceptance probability rate of 0.8 (this parameter controls the target acceptance rate of the NUTS algorithm, and setting it to a value close to 1 can reduce the number of

Table 1
Posterior predictive diagnostics for Case 1.

Metric	Partial pooling	No pooling			
	All blades	Blade 1	Blade 2	Blade 3	Blade 4
min N_{Eff}	3085	518	7583	1290	1251
max R_{hat}	1	1	1	1	1

false-positive divergences [54]). The sampler was run for 10 000 samples per chain (and an additional 5000 warm-up samples were discarded per chain to diminish the influence of the starting values [47]), for each parameter. Models were also run for each set of blade data independently (no pooling), for comparison with the partial-pooling model. Posterior predictive checks were performed to ensure stationarity and proper mixing of the Markov chains. As computed by Stan, the *maximum* split R_{hat} diagnostic, and the *minimum* effective sample size (N_{Eff}), considering all parameters/hyperparameters, are shown in Table 1, for all models. Note that split R_{hat} values that do not exceed 1.01 suggest proper mixing of the chains [55]. Likewise, total N_{Eff} values greater than 400 are considered suitable for stable estimates and proper interpretation of split R_{hat}, for a setup comprised of four parallel chains, as employed in this work [55]. These criteria were met for all models developed.

6.2. Model results

Eq. (13) was used to compute FRFs from the posterior MCMC samples of the modal parameters, for each blade. FRFs *with predicted variance* were generated by sampling from a Student's t-distribution with four degrees-of-freedom, mean equal to the computed FRFs, and posterior predictive measurement noise, (σ^2), for each sample. Total variance was estimated by taking the standard deviation of the FRFs. Posterior predictive mean and 3-sigma deviation for the partial-pooling and independent models are plotted in Figs. 9(a) to 10(b), respectively.

Indeed, Figs. 9(a) to 10(b) show that the sparser FRFs improved significantly for the combined model, compared to the independent models. With only 11 data points, Blade 3, shown in Fig. 10(a) in green, was missing most of the information necessary to characterise the two modes. The independent model for Blade 3 relied heavily on the user-specified prior, while the partial-pooling model borrowed information from the other three blades to inform the shared latent model. Similar (albeit less severe), results were seen for Blade 4, shown in Fig. 10(b) in purple. With 31 data points, the FRF for Blade 4 was less sparse than that for Blade 3, but was still somewhat lacking near the resonance peaks, especially for the first mode. Fig. 10(b) shows a significant reduction in variance with the partial-pooling approach. The improvements in the data-rich FRFs, Blades 1 (blue) and 2 (red) were not as apparent, as expected.

Variance reduction from the combined-inference (partial-pooling) approach can also be visualised by plotting the marginal distributions of the parameters/hyperparameters. For the population-level variables, marginal distributions were approximated by taking the posterior MCMC samples for expectation and variance for each variable (or shape parameters, for damping), and then drawing from a distribution (normal for natural frequency and residue, and beta for damping) with the same statistics as the parameter draws. Kernel density estimation (KDE) was then used to approximate marginal distributions using the population-level draws. (Note it is known what distribution a given sample from the posterior will take, e.g., expectation and variance are learnt assuming *a priori* a normally-distributed mode shape. However, the shape of the posterior distribution is itself unknown, i.e., the distribution over these distributions. Because the posterior is a random variable, KDE can be used as an approximation for the marginal distributions, using the posterior MCMC samples.) Marginal distributions for the domain-level variables were also approximated via KDE of the posterior MCMC samples. For the natural frequencies, the population- and domain-level distributions for each mode are shown in Figs. 11(a) and 11(b). For damping, the population- and domain-level distributions for each mode are shown in Figs. 11(c) and 11(d). For the (shared) residues, which were learnt at the population-level, the parent- and lower-level distributions for each mode are shown in Figs. 12(a) and 12(b).

Figs. 11(a) to 12(b) show that when the domains were able to share information via common higher-level distributions, the variance was reduced compared to a no-pooling approach. In these figures, the population-level distributions are shown at the top, while domain-level distributions are shown on the bottom of the plots. In Fig. 11(a), for the first natural frequency, the population-level distribution of the partial-pooling model was closely aligned with those of the independent models for the data-rich domains, and was taller and with lower variance than those of the independent models for the data-poor domains. This observation was likely because the data-rich domains dominated the available data for that mode, compared to the data-poor domains (forcing the distribution towards the natural frequencies of the data-rich domains). This data sharing was also evident at the domain-level, where the KDEs of the data-poor domains aligned closely with those from the data-rich domains, for the partial-pooling models. Also, in Fig. 11(a), it can be seen that the no-pooling KDEs of the data-poor FRFs were primarily informed by the priors. Indeed, the population-level distribution for the sparsest domain, Blade 3, was nearly indistinguishable from the prior, indicating that the data had little influence on the posterior (this makes sense, as there were very little data to characterise this mode). In contrast, Fig. 11(b) shows that for the second natural frequency, the population-level distribution for the partial-pooling model was flatter than those for the independent models. This was likely to have occurred because more data were available for the second peak from Blade 4, which was from the second grouping (widening the distribution to accommodate the greater differences in natural frequency). Again, this conclusion is supported at the domain-level, where the KDE for Blade 4 is very similar for the partial-pooling and independent models (although there is still a small variance reduction for the partial-pooling model). For Blade 3, which was
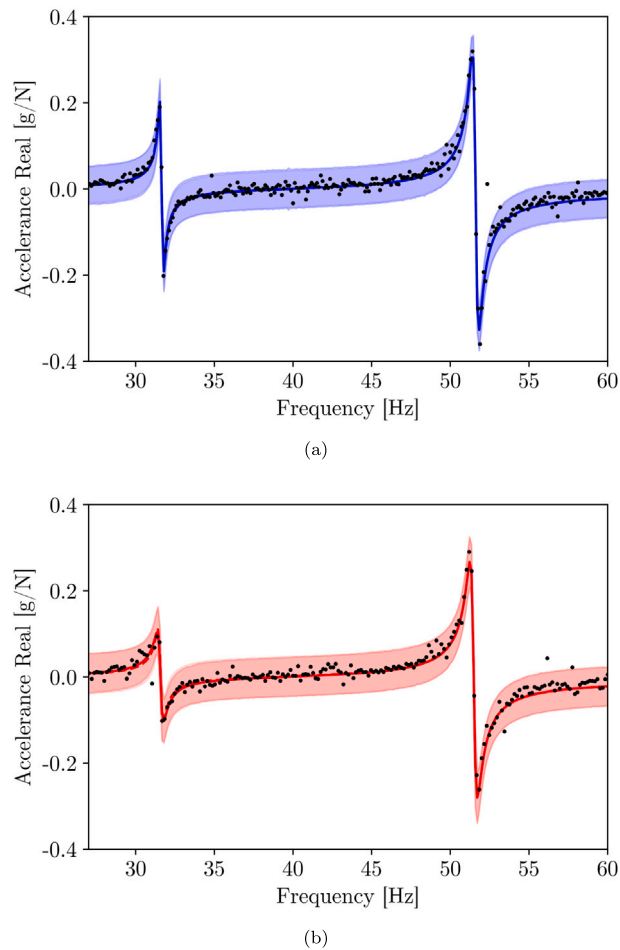
Fig. 9. Posterior predictive mean and 3-sigma deviation for independent (no-pooling) and partial-pooling models, for data-rich domains (a) Blade 1 (blue), and (b) Blade 2 (red). Training data are shown in a black scatter plot. Solid lines are the posterior predictive means for the partial-pooling models, and dashed lines are the posterior predictive means for the independent models. The variance is represented by shaded regions, where the independent model variances are shown in a lighter colour than those for the partial-pooling models.

quite data-poor at the second mode as well as the first, the KDE in Fig. 11(b) shows that the natural frequency was estimated to occur in between those of the other, more data-rich blades. Again, for Blade 3, the no-pooling KDE differed significantly from the others, as this parameter was primarily informed by the priors, because of a lack of data at the second mode. For both modes, it is clear that the independent models struggled to fit the data-poor FRFs, as evidenced by bimodal distributions for natural frequency.

Figs. 11(c) and 11(d) show that the population-level distributions for damping did not vary significantly among the partial-pooling and independent models (with only a slight deviation from the priors for the partial-pooling and data-rich models), likely because the priors were fairly strong and also accurate to the data. At the domain-level, KDEs for the data-rich FRFs showed high similarity for the different pooling models, while KDEs for the data-poor FRFs tended towards the priors. Similar results were seen for the residues, in Figs. 12(a) and 12(b), except that the residues were shared among the domains for the partial-pooling model, so there was only one lower-level distribution for each mode. Close alignment of the domain-level distributions, of the data-rich domains, suggests that sharing the residue among the different blades was a suitable assumption.

The results presented in this case demonstrate the development of a population form, where a combined hierarchical FRF model is learnt for a group of nominally-identical helicopter blades. Two domains (Blades 3 and 4) had limited data, and were especially sparse near the resonance peaks. Such a situation could occur if an insufficient sampling rate was selected during the data acquisition process. By borrowing data from data-rich domains within the population, variance was reduced compared to an independent modelling approach.

7. Case 2: population-based modelling of FRFs with temperature variation

The second case also used a hierarchical modelling structure with partial pooling, but assumed non-exchangeable models whereby parameters of the FRF varied with respect to temperature, i.e., the parameters were conditioned on the data. The goal was to
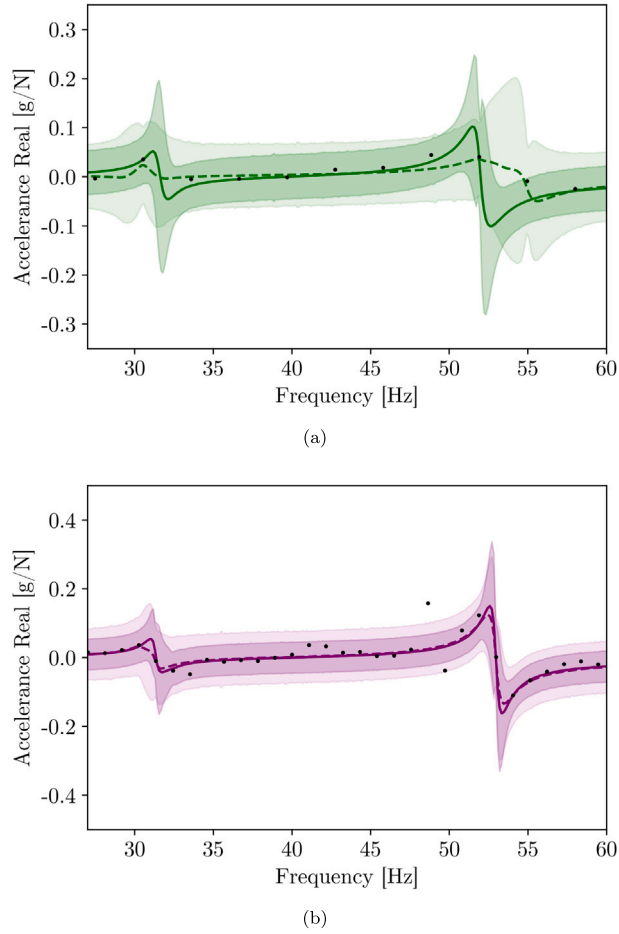
Fig. 10. Posterior predictive mean and 3-sigma deviation for independent (no-pooling) and partial-pooling models, for data-poor domains (a) Blade 3 (green) and (b) Blade 4 (purple). Training data are shown in a black scatter plot. Solid lines are the posterior predictive means for the partial-pooling models, and dashed lines are the posterior predictive means for the independent models. The variance is represented by shaded regions, where the independent model variances are shown in a lighter colour than those for the partial-pooling models.

learn functional relationships (at the population level), between temperature and natural frequency, and between temperature and damping, so that inferences could be made at temperatures not used in model training. As with the first case, all models in the second case were developed using `Stan`, and analyses were performed using MCMC, via the no U-turn (NUTS) implementation of Hamiltonian Monte Carlo (HMC) [51,52].

7.1. Identification of appropriate functional relationships between temperature and the modal parameters

Prior to developing the partial-pooling model, it was useful to determine appropriate functions for the temperature relationships. Independent models for each (real) temperature-varied FRF from Fig. 7(b) were fitted, and ground-truth values were estimated by computing the expectation for each modal parameter. (Note that in the second case, the modal parameters from the independent models for FRFs at all measured temperatures were considered to be the ground truth, as each FRF was data-rich. This process of using the same data to train and develop the model structure is called post-selection inference [56], and must be used cautiously, as models that employ the technique are at risk of overfitting [25,47]. However, it was assumed for this work that because the number of candidate models was small, the bias resulting from using the data twice was also small [25,47].) The ground-truth estimates for natural frequency and damping are plotted against temperature, and shown with least-squares fits (polynomial for natural frequency, and linear for damping), in Figs. 13(a) and 13(b). In the figures, the ground-truth estimates are plotted as asterisks, and those associated with the FRFs used to train the combined-inference model are shown in red, while those at 'unseen' temperatures are shown in black. The residue\mode shape was assumed to be constant with respect to temperature.

Fig. 13(a) shows that a second-order polynomial appears to be an appropriate fit for natural frequency in the temperature range of interest. From Fig. 13(b), a linear fit was considered most appropriate for damping. Although there may be a (weakly) nonlinear relationship between temperature and damping for this dataset, a linear assumption provided good results for the experiments
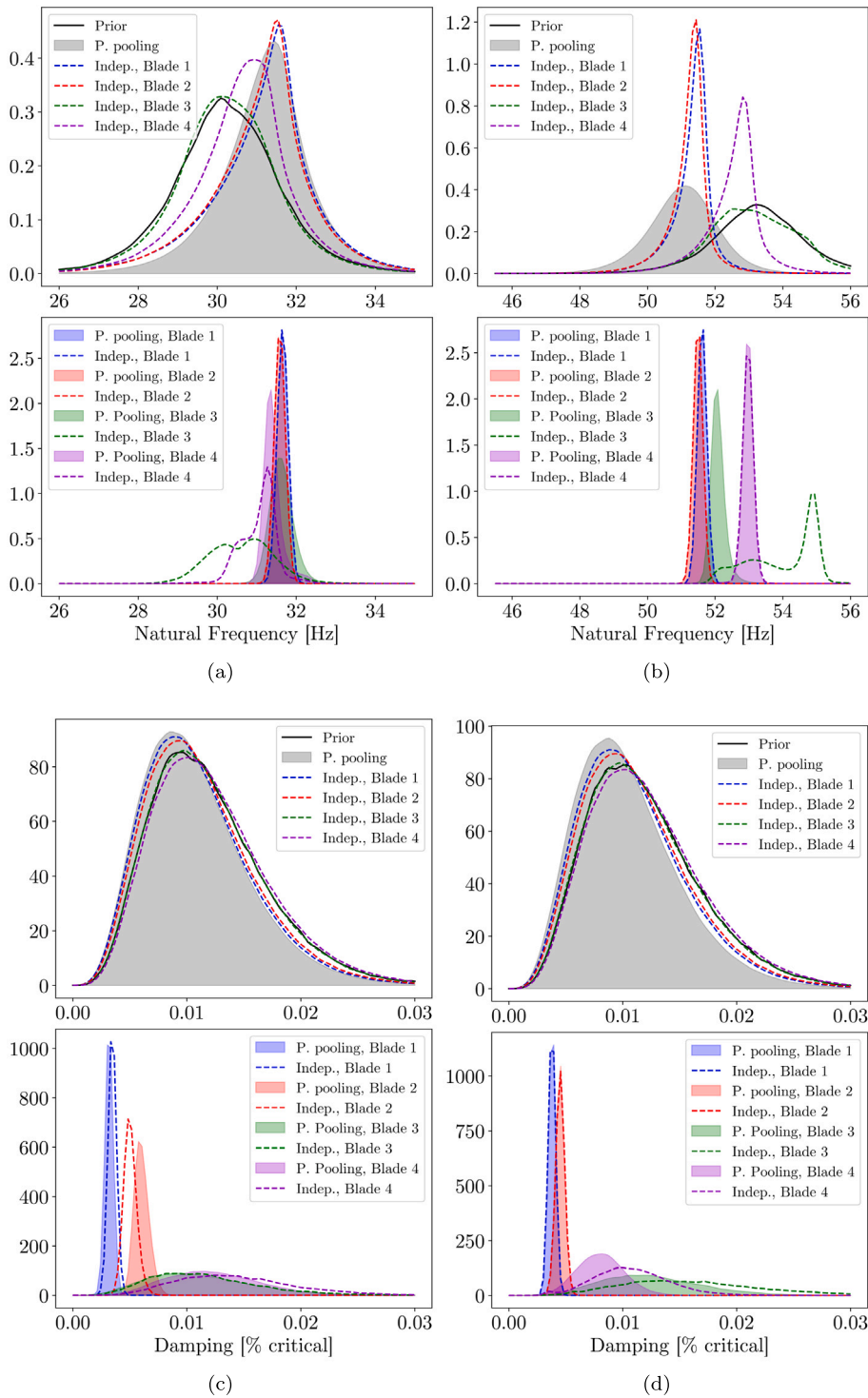
Fig. 11. Marginal distributions for natural frequency, (a) mode 1 and (b) mode 2; and for damping, (c) mode 1 and (d) mode 2. Population- and domain-level distributions are shown at the top and bottom, respectively. For the partial-pooling model, posterior distributions are shaded. For the independent (no-pooling) models, posterior distributions are shown as dashed lines. Priors are shown as solid black lines.
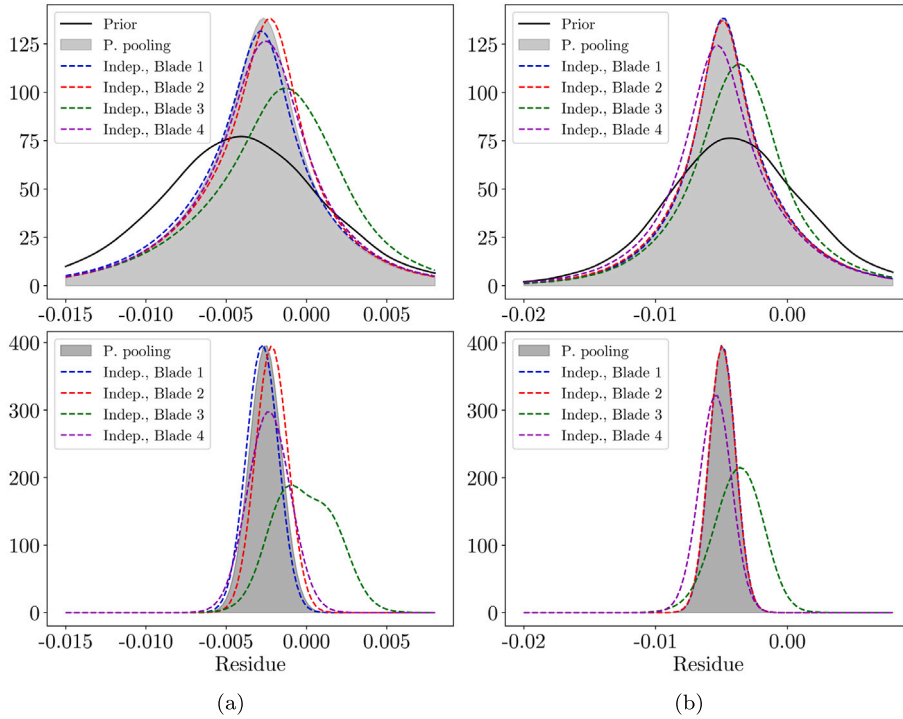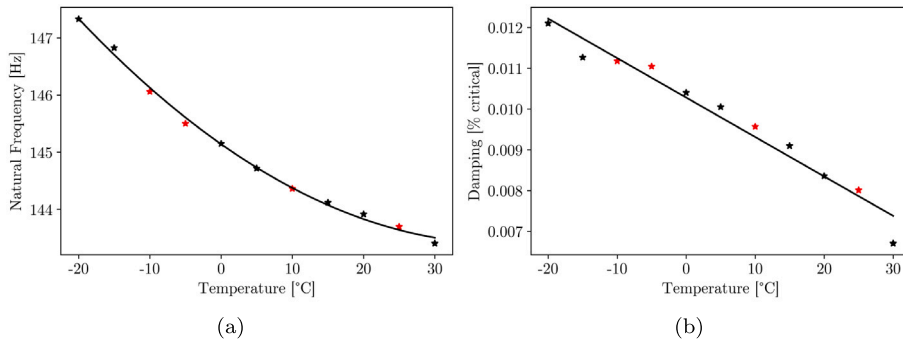
Fig. 12. Marginal distributions for residue, (a) mode 1 and (b) mode 2. Parent- and lower-level distributions are shown at the top and bottom, respectively. For the partial-pooling model, posterior distributions are shaded. For the independent (no-pooling) models, posterior distributions are shown as dashed lines. Priors are shown as solid black lines.

Fig. 13. Ground-truth estimates for (a) natural frequency and (b) damping, plotted against temperature. Second-order polynomial (for natural frequency) and linear (for damping) least-squares fits are shown as solid black lines. Parameters associated with FRFs used to train the model (−10, −5, 10, and 25 °C) are shown in red.

used here. A higher-order Taylor series expansion for the temperature-damping relationship would require many coefficients, which would significantly increase the number of hyperparameters in the model. Further studies, including material-properties evaluation, may help elucidate the nature of this relationship specific to the helicopter blades, which may allow for better inferences from increased physics-based knowledge.

7.2. Model development and results

As with the first case, the population model for the second case included four latent functions (i.e., $K = 4$), as four temperature-varied FRFs provided training data, i.e., those at −10, −5, 10, and 25 °C. In addition, the models developed were trained using data with added noise, to avoid over-fitting small deviations/lightly participating modes within the band of interest. This was accomplished by adding zero-mean Gaussian noise, with variance equal to 10% of the absolute maximum amplitude of the FRF obtained at −10 °C, to the 20 time-data blocks prior to computing FRFs. Then, the time data were subsampled such that all four

FRFs used in training were data rich (In other words, each FRF used in training had enough training points to fully characterise the mode in this band. Subsampling in this case was performed not to demonstrate shrinkage towards the data-rich groups, but to reduce the dimensionality of the training data). FRFs were then computed and averaged, and truncated to the desired bandwidth. This procedure resulted in 122 training points from each of the four temperature-varied FRFs. Also, once again, only the real part of the FRF was fitted for simplicity. (Again, given that the imaginary part could be modelled in the same fashion, or inferred from the real part, focussing on the real part only was deemed sufficient for demonstrating the techniques presented herein.) The real components of the FRFs were modelled probabilistically with an assumed Student's t-distribution, using the same accelerance FRF estimation as a likelihood function, as in Eqs. (13) and (18). (Note that as with the previous case, v was assumed to be four. This assumption was made for consistency with the previous case, although it could be learnt as a model parameter if desired.) The additive noise variance, σ^2, of the FRFs was assumed shared among each of the domains (again, the same data acquisition system and sensors were used among the different tests, and the same Gaussian noise were added), and the modal residue was assumed constant with respect to temperature.

Shared, higher-level distributions were placed over natural frequency, $\boldsymbol{\omega}_{nat} = \{\omega_{nat}^k\}_{k=1}^4$, damping, $\boldsymbol{\zeta} = \{\zeta_k\}_{k=1}^4$, and residue, A, to allow information sharing among the domains,

$$\mu_A \sim \mathcal{N}\left(\text{-0.008}, 0.002^2\right), \sigma_A^2 \sim \mathcal{TN}\left(0.002, 0.002^2\right)$$
$$\mu_{\omega_{nat}} \sim \mathcal{TN}\left(910, 10^2\right) \tag{23}$$
$$\mu_\zeta \sim \mathcal{TN}\left(0.01, 1^2\right)$$

Note that hyperpriors for $\boldsymbol{\omega}_{nat}$ are shown in rad/s. Also note that for this preliminary model, normal distributions were assumed for all parameters (with truncated normal distributions assumed for variance, damping, and natural frequency). The only exception is the noise variance, which was modelled as a half-Cauchy, as with the previous case. The residue, A, was assumed constant over all temperatures, with distributions,

$$A \sim \mathcal{N}\left(\mu_A, \sigma_A^2\right) \tag{24}$$

A second-order Taylor series expansion was used to approximate the functional relationship between temperature and natural frequency, with coefficients $\mathbf{a} = \{a_1, a_2\}$, where the domain-level natural frequencies were defined as temperature-shifted realisations from the population-level distributions, via,

$$\{\omega_{nat}^k = \mu_{\omega_{nat}} + a_1 T_k + a_2 T_k^2\}_{k=1}^4 \tag{25}$$

Likewise, the relationship between temperature and modal damping was approximated as linear, with slope b, over the measured temperatures, via,

$$\{\zeta_k = \mu_\zeta + b T_k\}_{k=1}^4 \tag{26}$$

where T_k was the temperature associated with the kth FRF. Priors were placed over the shared polynomial coefficients, with assumed distributions,

$$a_1 \sim \mathcal{N}\left(\text{-0.01}, 1^2\right)$$
$$a_2 \sim \mathcal{N}\left(0.001, 1^2\right) \tag{27}$$
$$b \sim \mathcal{N}\left(\text{-5e-6}, 1^2\right)$$

The shared noise variance, σ^2, was sampled from a half-Cauchy parent distribution, with higher-level scale parameter, γ^2, sampled from a truncated normal distribution,

$$\sigma^2 \sim C\left(0, \gamma^2\right)$$
$$\gamma^2 \sim \mathcal{TN}\left(0.01, 0.05^2\right) \tag{28}$$

A graphical model displaying the parameter hierarchy is shown in Fig. 14.

As stated previously, FRFs captured at temperatures −10, −5, 10, and 25 °C provided the training data for the model. The intent was to use a sub-set of the temperature-varied FRFs, to learn the population-level coefficients necessary to describe the relationships between temperature and the modal parameters, so that these population-level variables can be used to make predictions at 'unseen' temperatures (of course, these predictions were validated using experiments that did not contribute to the training data). As with the previous case, the Stan HMC sampler was run using four chains (and with a target average proposal acceptance probability rate of 0.8), for 10 000 samples per chain (with an additional 5000 warm-up samples per chain, which were discarded), for each parameter. Posterior predictive checks were performed to ensure stationarity and proper mixing of the Markov chains. As computed by Stan, the *maximum* split R_{hat} diagnostic, and the *minimum* effective sample size (N_{Eff}), considering all parameters/hyperparameters, are shown in Table 2, for all models (including those used for 'ground truth' natural frequency and damping estimates). Note that split R_{hat} values did not exceed 1.01, and N_{Eff} was greater than 400, suggesting proper mixing of the Markov chains, per current guidelines [55].

Eq. (13) was used to compute FRFs from the posterior MCMC samples of the modal parameters, for each temperature-varied FRF used in training. FRFs *with predicted variance* were generated by sampling from a Student's t-distribution with four degrees-of-freedom, mean equal to the computed FRFs, and posterior predictive measurement noise, σ^2, for each MCMC sample. Total variance
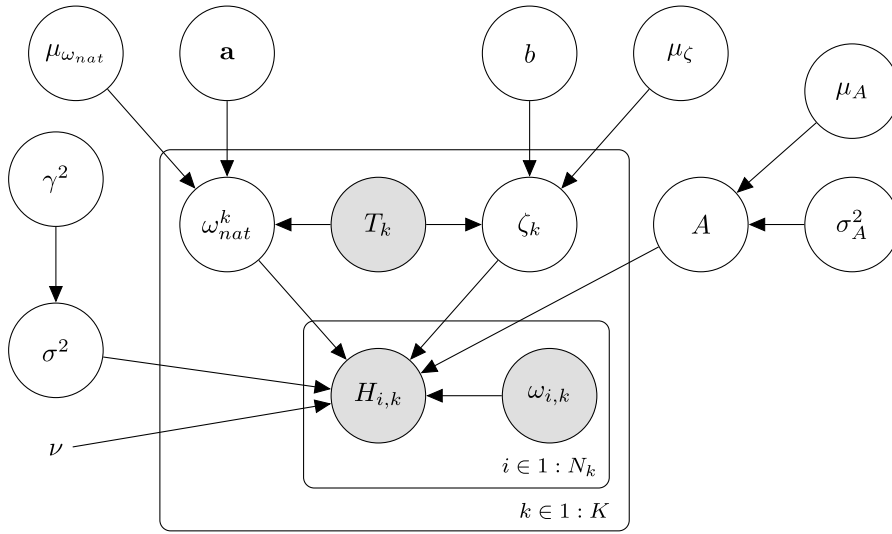
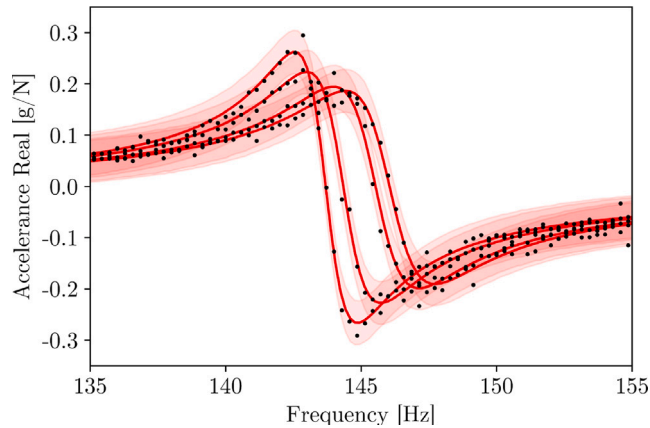**Fig. 14.** Graphical representation of hierarchical FRF model for the second case.

Fig. 15. Posterior predictive mean and 3-sigma deviation for probabilistic FRF model, for training data at −10, −5, 10, and 25 °C. Training data are shown in a black scatter plot. Solid lines are the posterior predictive means, and the variance is represented by shaded regions.

Table 2
Posterior predictive diagnostics for Case 2.

Metric	Partial-pooling model	'Ground-Truth' models
min N_{Eff}	15 000	640
max R_{hat}	1	1

Table 3
Expectation of the population-level variables.

E[A]	E[$\mu_{\omega_{nat}}$]	E[μ_ζ]	E[a_1]	E[a_2]	E[b]
−0.0083	145.1	0.0104	−0.5360	0.0075	−9.272e−5

was estimated by taking the standard deviation of the FRFs. Posterior predictive means and 3-sigma deviation for the FRFs are shown in Fig. 15.

The expectations for the population-level parameters were obtained from the samples, including the polynomial coefficients that define the relationships between temperature and the modal parameters. The expectations are shown in Table 3. The procedure for extrapolating beyond the FRFs used in training is described in Section 7.3.
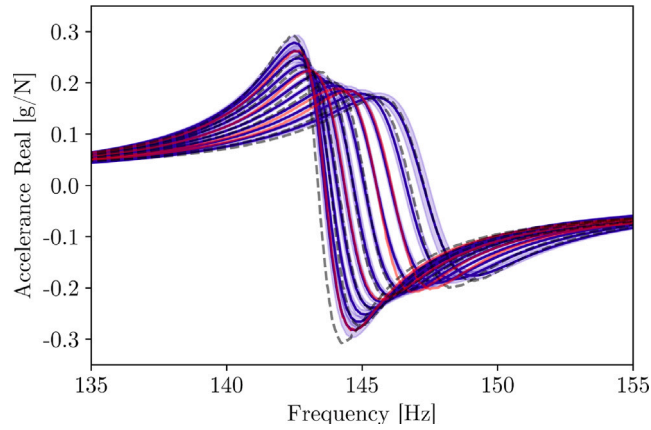
Fig. 16. Extrapolated FRFs using population-level variables, from −20 to 30 °C in increments of 5 °C, with decreasing temperature from left to right. The extrapolated FRF means are shown as solid blue lines, with shaded blue regions indicating variance bounds (3-sigma deviation). Solid red lines denote the FRFs that provided training data for the model, while dashed black lines denote the measured FRFs at 'unseen' temperatures.

Table 4
Evaluation of model accuracy for extrapolation to 'unseen' data.

Temperature [°C]	NMSE
−20	0.36
−15	0.86
−10	0.46
−5	0.59
0	0.68
5	0.52
10	0.42
15	0.47
20	0.67
25	0.80
30	3.36

7.3. Extrapolation to other temperatures

After training the model, predictions were made at 'unseen' temperatures, i.e., those not used in training. These predictions were made by computing natural frequency and damping using posterior MCMC samples, for all the measured temperatures, $T = \{T_i\}_{i=1}^{11}$, including the four used in training, via the relations,

$$\{\{\omega_{nat}^{i,n} = \mu_{\omega_{nat,n}} + a_{1,n}T_i + a_{2,n}T_i^2\}_{i=1}^{11}\}_{n=1}^{40000}$$
$$\{\{\zeta_{i,n} = \mu_{\zeta,n} + b_n T_i\}_{i=1}^{11}\}_{n=1}^{40000} \tag{29}$$

Natural frequency and damping were computed for each MCMC sample, for each temperature. These values were then used to compute FRFs via Eq. (13). Variance was estimated by taking the standard deviation of the FRFs. The extrapolated FRF means are plotted in Fig. 16, along with 3-sigma deviation. The extrapolated FRFs shown in Fig. 16 range from −20 to 30 °C in increments of 5 °C, with decreasing temperature from left to right, as natural frequency increases with decreasing temperature. Model accuracy was evaluated by calculating the NMSE via Eq. (15) for each measured FRF without added noise, compared to the mean extrapolated FRFs, as shown in Table 4. Note that the NMSE computed for the prediction at temperatures −10, −5, 10, and 25 °C, which correspond to the data used in model training, are included in the table for completeness. However, the specific training points were excluded from the calculation. To further visualise the relationships between temperature and the modal parameters, the functions were plotted for each sample. The mean and variance of the functions were plotted with the parameters estimated via independent modelling (in this case, these results were considered to be the best approximation of the ground truth, as each FRF was data-rich). These plots are shown in Figs. 17(a) and 17(b).

In Fig. 16, the extrapolated FRFs are shown as solid blue lines, with a shaded blue region indicating the variance bounds. The measured FRFs used to train the model, without added noise, are shown as solid red lines, and the measured FRFs used to test the model at 'unseen' temperatures are shown as dashed black lines. From the figure, it is clear that excellent agreement was achieved between the extrapolated and measured FRFs at the temperatures used to train the model, as expected. Likewise, the FRFs at temperatures between those used to train the model (e.g., at 0, 15, and 20 °C) show excellent agreement. Notably, the FRFs at colder temperatures, which were further away from the training data, still show good agreement. The accuracy of the fit is further shown in Table 4, as all NMSE values were less than 5%.
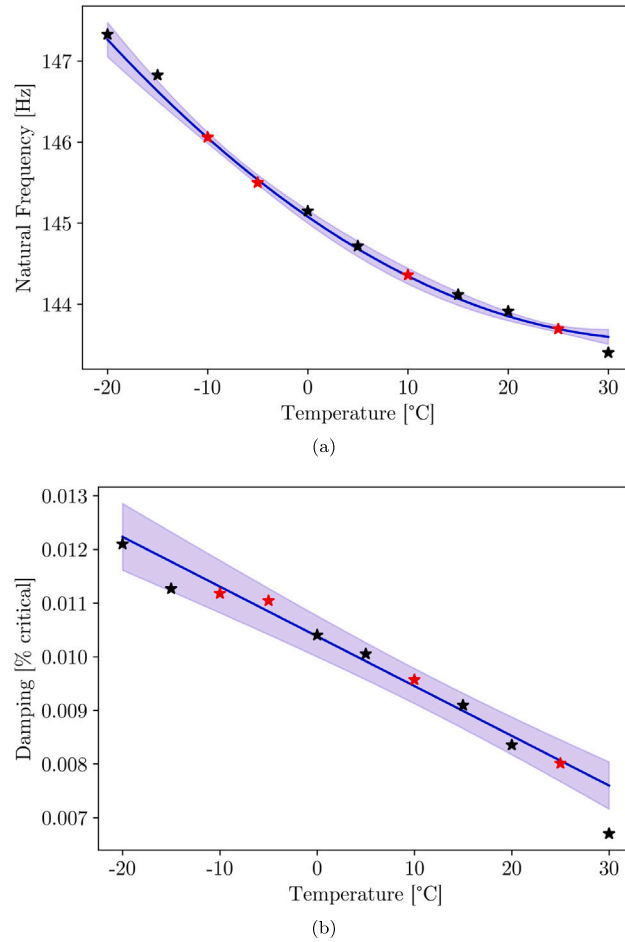
Fig. 17. Mean extrapolated (a) natural frequency and (b) damping for Blade 1, at training and 'unseen' temperatures, plotted against temperature, with 3-sigma deviation (shown in blue). Ground-truth model variables are shown as stars, with red indicating the temperatures used in training and black denoting 'unseen' temperatures.

Figs. 17(a) and 17(b) further show that the extrapolated parameters accurately represented the measured data, as the measured parameters largely fell within the variance bounds of the approximations. Fig. 17(a) shows that for natural frequency, parameter variance increased as the inferred parameters varied further from the training data (i.e., at temperatures further from those used to train the model), as expected. Fig. 17(b) shows that for damping, the relatively consistent variance was the result of the linear assumption for the temperature-damping relationship, and the relatively high deviation of the measured data from the linear fit. Further model development could involve the incorporation of more physics-based knowledge, such as, forcing the relationship between natural frequency and temperature to be monotonically decreasing, or further investigation into the possibly nonlinear relationship between damping and temperature. In addition, it may be possible to augment the model by imposing a sparsity-inducing prior on the weights of the function describing the relationship. This sparsity-inducing prior would allow the Bayesian inference to perform system identification with automatic relevance detection, allowing incorporation of additional terms, polynomial or otherwise.

8. Concluding remarks

Current work has involved the development of probabilistic FRF models using a hierarchical Bayesian approach, that account for benign variations as well as similarities among nominally-identical structures. Two cases were presented, to demonstrate the usefulness of this modelling structure. The first case demonstrated how hierarchical Bayesian models with partial pooling can reduce variance in data-poor domains by allowing information transfer with data-rich domains, via shared population-level distributions. The second case showed that incorporating functional relationships (approximated via Taylor series expansion) into the modelling structure to describe temperature variations allows for prediction beyond the training data. The first case addresses data-sparsity challenges in SHM, as missing time-domain data resulting from sensor dropout and other causes will reduce the number of spectral
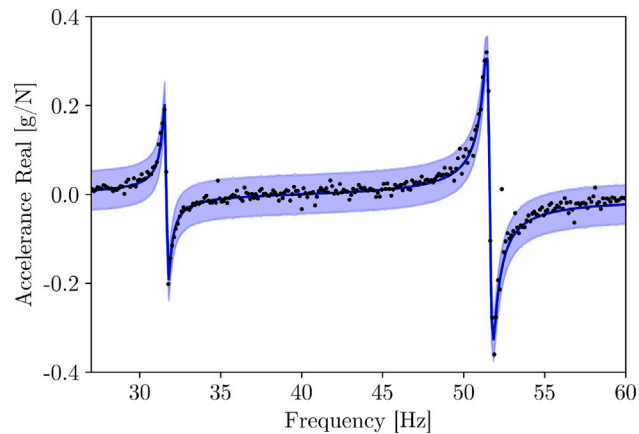
Fig. A.1. Posterior predictive mean and 3-sigma deviation for independent (no-pooling) and partial-pooling models, for data-rich Blade 1. Training data are shown in a black scatter plot. The posterior predictive means for the partial-pooling and independent models are shown as solid and dashed lines, respectively. The variance is represented by shaded regions, where the independent model variance is shown in a lighter colour than that for the partial-pooling model.

lines in the frequency domain, which can impede dynamic characterisation. Likewise, the second case addresses data-scarcity issues, as encoding physics-based knowledge into the model via functional relationships can increase the amount of normal-condition information available. Future efforts may involve further investigation into the physical relationships between temperature and the modal parameters, particularly damping, to improve model accuracy.

CRediT authorship contribution statement

T.A. Dardeno: Conceptualization, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **K. Worden:** Conceptualization, Funding acquisition, Methodology, Resources, Validation, Writing – review & editing. **N. Dervilis:** Conceptualization, Methodology, Resources, Validation, Writing – review & editing. **R.S. Mills:** Investigation, Resources, Validation. **L.A. Bull:** Conceptualization, Methodology, Software, Validation, Writing – review & editing.

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

Data will be made available on request.

Appendix. One data-rich FRF, three data-poor FRFs

Consider a situation, where comprehensive data were only known for a single structure in the population, with sparse data available for the remaining three structures. As with the case where two structures were data rich, and two were data poor, the result was shrinkage towards the population mean. The population mean was largely informed by the data-rich group, and prior information, with varying contributions from the sparse groups depending on how informative the available data were. Posterior predictive mean and 3-sigma deviation for the partial-pooling and independent models are plotted in Figs. A.1 to A.4, respectively. Note that the variance reduction was especially significant for Blade 2, as the data for this blade were quite uninformative.
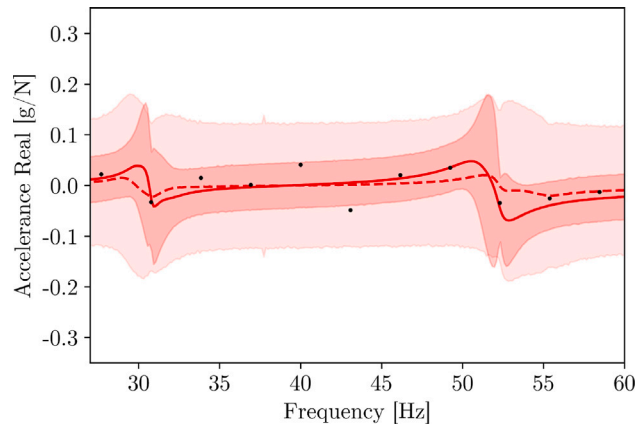
Fig. A.2. Posterior predictive mean and 3-sigma deviation for independent (no-pooling) and partial-pooling models, for data-poor Blade 2. Training data are shown in a black scatter plot. The posterior predictive means for the partial-pooling and independent models are shown as solid and dashed lines, respectively. The variance is represented by shaded regions, where the independent model variance is shown in a lighter colour than that for the partial-pooling model.

Fig. A.3. Posterior predictive mean and 3-sigma deviation for independent (no-pooling) and partial-pooling models, for data-poor Blade 3. Training data are shown in a black scatter plot. The posterior predictive means for the partial-pooling and independent models are shown as solid and dashed lines, respectively. The variance is represented by shaded regions, where the independent model variance is shown in a lighter colour than that for the partial-pooling model.
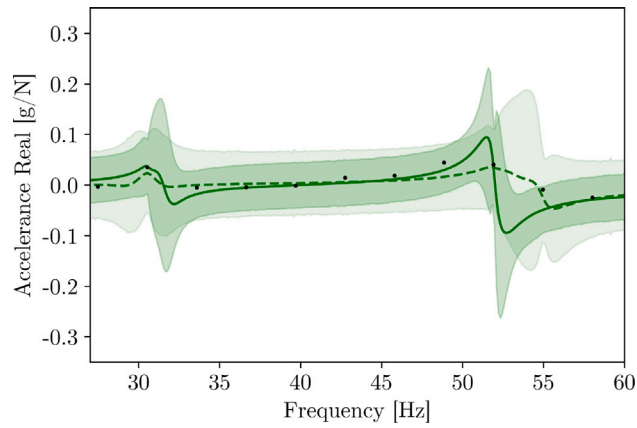

Fig. A.4. Posterior predictive mean and 3-sigma deviation for independent (no-pooling) and partial-pooling models, for data-poor Blade 4. Training data are shown in a black scatter plot. The posterior predictive means for the partial-pooling and independent models are shown as solid and dashed lines, respectively. The variance is represented by shaded regions, where the independent model variance is shown in a lighter colour than that for the partial-pooling model.
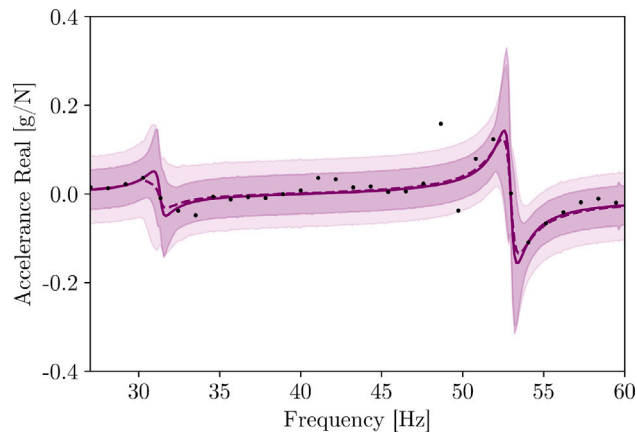
References

[1] P. Cawley, A.M. Woolfrey, R.D. Adams, Natural frequency measurements for production quality control of fibre composites, Composites 16 (1985) 23–27.

[2] W.D. Zhu, K. He, Detection of damage in space frame structures with L-shaped beams and bolted joints using changes in natural frequencies, J. Vib. Acoust. 135 (2013).

[3] M. Colakoglu, Effect of temperature on frequency and damping properties of polymer matrix composites, Adv. Compos. Mater. 17 (2008) 111–124.

[4] H. Watanabe, T. Mukai, M. Sugioka, K. Ishikawa, Elastic and damping properties from room temperature to 673 K in an AZ31 magnesium alloy, Scr. Mater. 51 (2004) 291–295.

[5] Y. Bai, K. Yu, J. Zhao, R. Zhao, Experimental and simulation investigation of temperature effects on modal characteristics of composite honeycomb structure, Compos. Struct. 201 (2018) 816–827.

[6] K. Worden, H. Sohn, C. Farrar, Novelty detection in a changing environment: Regression and interpolation approaches, J. Sound Vib. 258 (2002) 741–761.

[7] H. Sohn, Effects of environmental and operational variability on structural health monitoring, Philos. Trans. A: Math. Phys. Eng. Sci. 365 (2007) 539–560.

[8] S. Alampalli, Influence of in-service environment on modal parameters, in: Proceedings of the 16th International Modal Analysis Conference (IMAC), Society for Experimental Mechanics, Santa Barbara, CA, USA, 1998.

[9] P. Cawley, Long range inspection of structures using low frequency ultrasound, in: Proceedings of the International Conference on Damage Assessment of Structures (DAMAS 97), Sheffield Academic Press, Sheffield, UK, 1997, pp. 1–17.

[10] M. Cao, G. Sha, Y. Gao, W. Ostachowicz, Structural damage identification using damping: A compendium of uses and features, Smart Mater. Struct. 26 (2017) 043001.

[11] J. Pei, C. Kapoor, T. Graves-Abe, Y. Sugeng, J. Lynch, Critical design parameters and operating conditions of wireless sensor units for structural health monitoring, in: Proceedings of the 23rd International Modal Analysis Conference (IMAC), 2005.

[12] J. Meyer, R. Bischoff, G. Feltrin, M. Motavalli, Wireless sensor networks for long-term structural health monitoring, Smart Struct. Syst. 6 (2010).

[13] Y.Y. Bao, H. Li, X. Sun, Y. Yu, J. Ou, Compressive sampling–based data loss recovery for wireless sensor networks used in civil structural health monitoring, Struct. Health Monit. 12 (2013) 78–95.

[14] A. Malekloo, E. Ozer, M. AlHamaydeh, M. Girolami, Machine learning and structural health monitoring overview with emerging technology and high-dimensional data source highlights, Struct. Health Monit. 21 (2022) 1906–1955.

[15] J. Lynch, K. Loh, A summary review of wireless sensors and sensor networks for structural health monitoring, Shock Vib. Digit. 38 (2006) 91–128.

[16] T. Nagayama, S. Sim, Y. Miyamori, B.F. Spencer Jr., Issues in structural health monitoring employing smart sensors, Smart Struct. Syst. 3 (2007) 299–320.

[17] J. Gosliga, P.A. Gardner, L.A. Bull, N. Dervilis, K. Worden, Foundations of population-based SHM, Part II: Heterogeneous populations–graphs, networks, and communities, Mech. Syst. Signal Process. 148 (2021) 107144.

[18] P.A. Gardner, L.A. Bull, J. Gosliga, N. Dervilis, K. Worden, Foundations of population-based SHM, Part III: Heterogeneous populations–mapping and transfer, Mech. Syst. Signal Process. 149 (2021) 107142.

[19] L.A. Bull, P.A. Gardner, J. Gosliga, T.J. Rogers, N. Dervilis, E.J. Cross, E. Papatheou, A.E. Maquire, C. Campos, K. Worden, Foundations of population-based SHM, Part I: Homogeneous populations and forms, Mech. Syst. Signal Process. 148 (2021) 107141.

[20] M. Lázaro-Gredilla, S. Van Vaerenbergh, N.D. Lawrence, Overlapping mixtures of Gaussian processes for the data association problem, Pattern Recognit. 45 (2012) 1386–1395.

[21] L.A. Bull, P.A. Gardner, T.J. Rogers, N. Dervilis, E.J. Cross, E. Papatheou, A.E. Maquire, C. Campos, K. Worden, Bayesian modelling of multivalued power curves from an operational wind farm, Mech. Syst. Signal Process. (2021) 108530.

[22] T.A. Dardeno, L.A. Bull, R.S. Mills, N. Dervilis, K. Worden, Modelling variability in vibration-based PBSHM via a generalised population form, J. Sound Vib. 538 (2022) 117227.

[23] L.A. Bull, M. Dhada, O. Steinert, T. Lindgren, A. Kumar, M. Girolami, Population-level modelling for truck fleet survival analysis, in: Proceedings of the 10th European Workshop on Structural Health Monitoring (EWSHM), Springer, Palermo, Italy, 2022.

[24] L.A. Bull, D. Di Francesco, M. Dhada, O. Steinert, T. Lindgren, A.K. Parlikad, A.B. Duncan, M. Girolami, Hierarchical Bayesian modeling for knowledge transfer across engineering fleets via multitask learning, Comput.-Aided Civ. Inf. 38 (2023) 821–848.

[25] L.A. Bull, M.R. Jones, E.J. Cross, A. Duncan, M. Girolami, Encoding domain expertise into multilevel models for source location, 2023, arXiv preprint arXiv:2305.08657.

[26] 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.

[27] R-T. Wu, A. Singla, M.R. Jahanshahi, E. Bertino, B.J. Ko, D. Verma, Pruning deep convolutional neural networks for efficient edge computing in condition assessment of infrastructures, Comput.-Aided Civ. Inf. 34 (2019) 774–789.

[28] S. Dorafshan, R.J. Thomas, M. Maguire, Comparison of deep convolutional neural networks and edge detectors for image-based crack detection in concrete, Constr. Build. Mater. 186 (2018) 1031–1045.

[29] Y. Gao, K.M. Mosalam, Deep transfer learning for image-based structural damage recognition, Comput.-Aided Civ. Inf. 33 (2018) 748–768.

[30] K. Jang, N. Kim, Y-K. An, Deep learning–based autonomous concrete crack evaluation through hybrid image scanning, Struct. Health Monit. 18 (2019) 1722–1737.

[31] H. Perez, J.H.M. Tah, A. Mosavi, Deep learning for detecting building defects using convolutional neural networks, Sensors 19 (2019).

[32] G. Michau, O. Fink, Domain adaptation for one-class classification: Monitoring the health of critical systems under limited information, Int. J. Progn. Health Manage. 10 (2019) 4.

[33] A. Ozdagli, X. Koutsoukos, Domain adaptation for structural health monitoring, in: Annual Conference of the PHM Society, Vol. 12, 2020, p. 9.

[34] Z. Sun, A. Barp, F-X. Briol, Vector-valued control variates, 2021, arXiv preprint arXiv:2109.08944.

[35] H-P. Wan, Y-Q. Ni, Bayesian multi-task learning methodology for reconstruction of structural health monitoring data, Struct. Health Monit. 18 (2019) 1282–1309.

[36] Y. Li, T. Bao, Z. Chen, Z. Gao, X. Shu, K. Zhang, A missing sensor measurement data reconstruction framework powered by multi-task Gaussian process regression for dam structural health monitoring systems, Measurement 186 (2021) 110085.

[37] P. Seshadri, A.B. Duncan, G. Thorne, G. Parks, R.V. Diaz, M. Girolami, Bayesian assessments of aeroengine performance with transfer learning, Data-Centric Eng. 3 (2022) e29.

[38] G.C. Ballesteros, P. Angelikopoulos, C. Papadimitriou, P. Koumoutsakos, Bayesian hierarchical models for uncertainty quantification in structural dynamics, in: Vulnerability, Uncertainty, and Risk: Quantification, Mitigation, and Management, 2014, pp. 1615–1624.

[39] I. Behmanesh, B. Moaveni, G. Lombaert, C. Papadimitriou, Hierarchical Bayesian model updating for structural identification, Mech. Syst. Signal Process. 64 (2015) 360–376.

[40] Y. Huang, J.L. Beck, Hierarchical sparse Bayesian learning for structural health monitoring with incomplete modal data, Int. J. Uncertain. Quantif. 5 (2015).

[41] Y. Huang, J.L. Beck, H. Li, Multitask sparse Bayesian learning with applications in structural health monitoring, Comput.-Aided Civ. Inf. 34 (2019) 732–754.

[42] A. Poblete, R.O. Ruiz, G. Jia, Hierarchical Bayesian approach for model parameter updating in piezoelectric energy harvesters, Mech. Syst. Signal Process. 172 (2022) 108942.

[43] D. Di Francesco, M. Chryssanthopoulos, M.H. Faber, U. Bharadwaj, Decision-theoretic inspection planning using imperfect and incomplete data, Data-Centric Eng. 2 (2021).

[44] N. Papadimas, T. Dodwell, A hierarchical Bayesian approach for calibration of stochastic material models, Data-Centric Eng. 2 (2021).

[45] M. Dhada, M. Girolami, A.K. Parlikad, Anomaly detection in a fleet of industrial assets with hierarchical statistical modeling, Data-Centric Eng. 1 (2020) e21.

[46] O. Sedehi, A.M. Kosikova, C. Papadimitriou, L.S. Katafygiotis, On the integration of physics-based machine learning with hierarchical Bayesian modeling techniques, 2023, arXiv preprint arXiv:2303.00187.

[47] A. Gelman, J.B. Carlin, H.S. Stern, D.B. Dunson, A. Vehtari, D.B. Rubin, Bayesian Data Analysis, CRC Press, Boca Raton, FL, 2013.

[48] K. Worden, G.R. Tomlinson, Nonlinearity in Structural Dynamics: Detection, Identification and Modelling, Institute of Physics Publishing, Bristol, UK, 2001.

[49] D. Garinis, M. Dinulović, B. Rašuo, Dynamic analysis of modified composite helicopter blade, FME Trans. 40 (2012) 63–68.

[50] T.A. Dardeno, R.S. Mills, L.A. Bull, N. Dervilis, K. Worden, Investigating the effects of temperature variation in vibration-based SHM, in: Proceedings of the 30th International Conference on Noise and Vibration Engineering (ISMA), Springer, Leuven, Belgium, 2022.

[51] M.D. Hoffman, A. Gelman, The No-U-Turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo, J. Mach. Learn. Res. 15 (2014) 1593–1623.

[52] M. Betancourt, M. Girolami, Hamiltonian Monte Carlo for hierarchical models, Curr. Trends Bayes. Methodol. Appl. 79 (2015) 2–4.

[53] A. Gelman, Prior distributions for variance parameters in hierarchical models 1 (2006) 515–533.

[54] M. Betancourt, Diagnosing biased inference with divergences, 2017, https://betanalpha.github.io/assets/case_studies/divergences_and_bias.html, Last accessed on May 27 2023.

[55] A. Vehtari, A. Gelman, D. Simpson, B. Carpenter, P.C. Bürkner, Rank normalization, folding, and localization: An improved $\hat{r}$ for assessing convergence of MCMC (with discussion), Bayesian Anal. 16 (2021) 667–718.

[56] J.D. Lee, D.L. Sun, Y. Sun, J.E. Taylor, Exact post-selection inference, with application to the lasso, Ann. Statist. 44 (2016) 907–927.