

Towards Multilevel Modelling of Train Passing Events on the Staffordshire Bridge

LAWRENCE A. BULL¹, CHIHO JEON¹, MARK GIROLAMI^{1,4},
ANDREW DUNCAN^{2,4}, JENNIFER SCHOOLING¹
and MIGUEL BRAVO HARO³

ABSTRACT

We suggest a multilevel model, to represent aggregate train-passing events from the Staffordshire bridge monitoring system. We formulate a combined model from simple units, representing strain envelopes (of each train passing) for two types of commuter train. The measurements are treated as a longitudinal dataset and represented with a (low-rank approximation) hierarchical Gaussian process. For each unit in the combined model, we encode domain expertise as boundary condition constraints and work towards a general representation of the strain response. Looking forward, this should allow for the simulation of train types that were previously unobserved in the training data. For example, trains with more passengers or freights with a heavier payload. The strain event simulations are valuable since they can inform further experiments (including FEM calibration, fatigue analysis, or design) to test the bridge in hypothesised scenarios.

INTRODUCTION

An increasing number of bridge-monitoring projects utilise streaming telemetry data [1–4]. It is vital that we develop appropriate statistical models to represent and extract valuable insights from these large datasets, since these bridges constitute critical infrastructure within modern transportation networks. The process of monitoring engineered systems via streaming data is typically referred to as Structural Health Monitoring (SHM) and while successful applications have been emerging in recent years, a number of challenges remain for practical implementation [5]. During model design, these concerns usually centre around *low variance* data: that is, measurements are not available for the entire range of expected operational, environmental, and damage conditions. Consider a bridge following construction, this will have a relatively small dataset that should only be associated with *normal* operation. On the other hand, a structure

Lawrence Bull, lab99@cam.ac.uk. ¹University of Cambridge, Department of Engineering, Cambridge, CB3 0FA, UK; ²Imperial College London, Department of Mathematics, London, SW7 2AZ, UK; ³City, University of London, Department of Engineering, London, EC1V 0HB, UK; ⁴The Alan Turing Institute, The British Library, 96 Euston Road, London, NW1 2DB, UK

with historical data might still not experience low-probability events – such as extreme weather or landslides.

An obvious solution considers sharing data (or information) between structures; this has been the focus of a large body of recent work [6–8]. Since no two structures are identical, simply combining the data (complete pooling) is rarely sufficient¹. A number of statistical and ML approaches become suitable in this setting: domain adaptation [8], transfer learning [10], multitask learning [11], and federated learning [12] are a few, recently explored examples.

In this article, we present a model design for data from the landmark IB5 (Staffordshire) rail bridge digital twin [2, 13]. A convenient artefact of rail bridge monitoring is the repeated *and comparable* train passing events, which readily allow information to be shared in ways that are less feasible for foot or automobile bridges. Train events occur frequently, with an average of 60-70 trains crossing IB5 per day. Many of these correspond to passenger/commuter trains, the focus of this initial work. We propose a Gaussian process representation, to work towards encoding prior knowledge of train events within a general, flexible representation.

The layout of this paper is as follows. Section 3 outlines the existing IB5 sensing system; Section 4 summarises a decision tree to classify train events; Section 5 formulates the multilevel Gaussian process representation; Section 6 presents initial monitoring insights from the model; Section 7 offers concluding remarks and summarises future work.

THE NORTON (IB5) BRIDGE

It is widely accepted that monitoring and analysing telemetry data from bridges provides valuable insights. Though in practice, there are few examples of permanent and reliable sensing systems, which allow for the remote acquisition of *streaming data*. The Norton intersection bridge 5 (IB5) is one example, with a sensing system that continuously monitors signals, such that train events are recorded, stored, and made accessible via an API [14].

The IB5 structure itself is a half-through, E-type bridge owned by Network Rail, with a single skew axis span of 26.8 m, shown in Figure 1. It carries two rail lines on the West Coast Main Line near Crewe, UK, constructed as part of a rail redevelopment project – the Stafford Area Improvement Programme.

The bridge has been instrumented with numerous sensors. Fibre Bragg Gratings (FBGs) were installed during construction. These were attached to the main and cross beams, and embedded in the reinforced concrete deck. More recently, accelerometers were deployed throughout the main load-bearing steel girder, as well as optical, humidity, and temperature sensors (and video cameras) [15].

The acquisition system monitors all sensors in real-time, capturing a synchronised batch of raw measurements with each train passing. Over 34,000 passing events have been captured at the time of writing.

¹Although, in certain cases, complete pooling works, especially for more simple operations such as novelty detection with consistent data [6, 9]



Figure 1. The IB5 Norton bridge, a composite structure with steel I-girders and a *cast-in-situ* reinforced concrete deck, spanning 26.8m

Strain and BWIM Data

This work considers strain data and features from a Bridge Weigh in Motion (BWIM) system, implemented in previous phases of the IB5 project [16]. FBG raw measurements are *changes in wavelength* $\Delta\lambda$, which we consider from the west main girder of the bridge only. This is converted to the *change in micro-strain* ($\Delta\mu\varepsilon$) as follows [13],

$$\Delta\mu\varepsilon = \frac{1}{k_\varepsilon} \cdot \frac{\Delta\lambda}{\lambda_0} \times 10^{-6} \quad (1)$$

where k_ε is the gauge factor provided by the FBG manufacturer (0.78), and λ_0 scales raw measurements ($\Delta\lambda$) to present a relative change. Since we hope to compare the bridge response for many events, the reference λ_0 is an average from the start of each time series. In practical terms, this scales each response to zero at the start of the strain envelope. Also, note that FBG sensors are highly sensitive to temperature variations; however, here we assume that changes in temperature are negligible for each event, and leave out the associated compensation (which can be found in [13]).

Figure 2 presents an example of the strain data over time. The underlying function has a convenient interpretation: the sum of *strain influence lines* per axle, which can be used to estimate each axle weight contribution [17]. Currently, strain influence lines are only utilised within the BWIM system, to classify passing events according to train type (Section 4). However, looking forward, parametrising the function as a sum over (influence line) bases presents an interesting modelling perspective – discussed as future work in Sections 5 and 7.

Conceptually, the BWIM system turns the bridge into a weighing scale, estimating each axle weight from *in-situ* monitoring data. The result is a number of insightful features, three are considered in this work: (i) number of axles, (ii) axle weights, (iii) axle spacing, and transformations thereof.

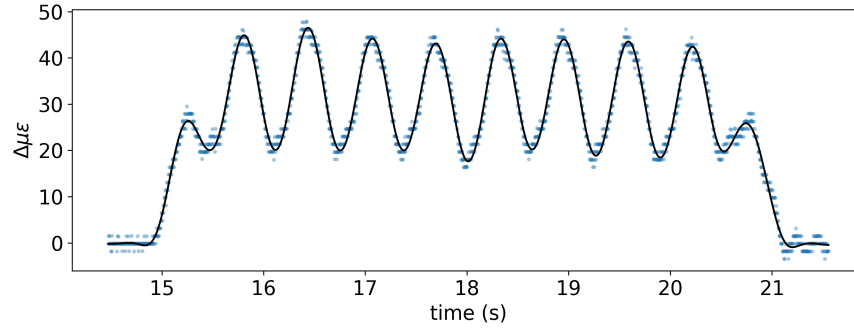


Figure 2. Strain time-series data (blue) and the filtered signal (black).

TRAIN EVENT CLASSIFICATION

The focus of this article is to design a general representation of train events. With this in mind, it makes sense to initially consider *typical* strain responses, and then alter the representation, if required (ideally without increasing model complexity). An obvious starting point is 16-axle commuter train types; including 350, 220, and 221 trains [13]. Given the BWIM features extracted from each train passing, alongside knowledge of train design, a simple decision tree can be used to isolate the commuter train data. The categorisation is presented in Figure 3.

In the current work, we consider both 350 and 22x trains, highlighted with **green** and **purple** text at the leaves of the tree in Figure 3. An example of ten 350 passing events is shown in Figure 4. Note the input normalisation, using passing distance rather than time, which horizontally scales each strain signal to be comparable, irrespective of the train-speed.

Observing Figure 4, it becomes clear that the strain data are characteristic and repeatable (in this case, for 350 trains). In turn, their underlying function should be ideal for both monitoring and finding a general representation. Here we explore a basic Gaussian process representation, which captures the characteristic response, alongside an appropriate uncertainty quantification for the longitudinal data. Before inference, we z-score normalise the inputs and divide the response by its scale, as shown in Figure 5. This allows for stable inferences via Hamiltonian Monte Carlo and convenient prior definitions in the following section.

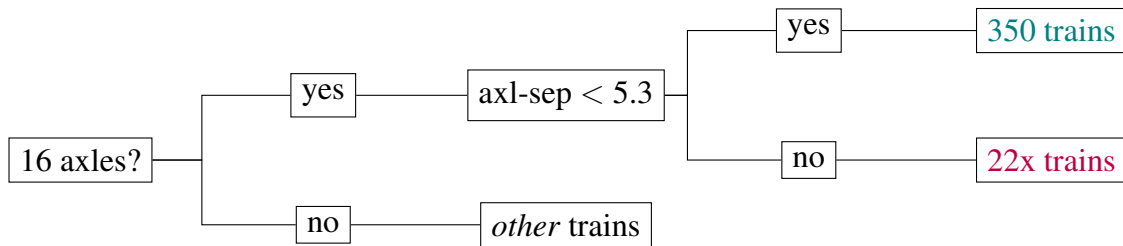


Figure 3. Train classification using BWIM data: [axl-sep] is the average axle separation, and 22x refers to either 220 or 221 trains.

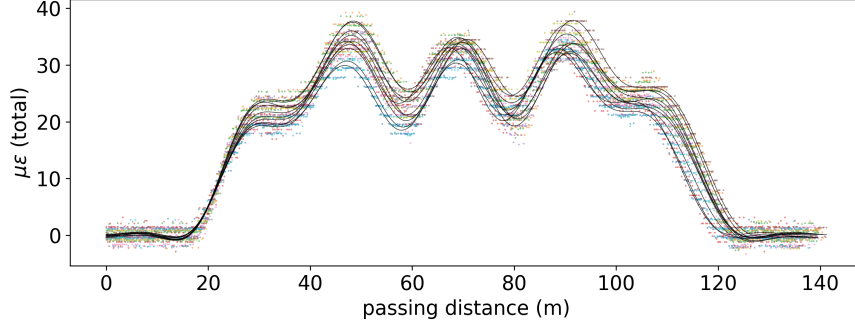


Figure 4. An example of ten passing events for 350 trains. Different colour markers for each train.

MODEL DESIGN

We now describe the multilevel Gaussian process to represent the aggregated data. The multilevel structure will naturally extend to trains of the same type, while also presenting options to incorporate different designs. Importantly, with some (mean function) modifications, the model should represent the uncertainty of *event-specific* functions, alongside the uncertainty due to variations between different events of the same train type, i.e. *type-specific* functions.

We use the notation \mathbf{y}_k to denote the strain series (response vector) for event $k \in \{1, 2, \dots, K\}$ with the corresponding inputs \mathbf{x}_k (passing distance). Each event is then approximated as some function f_k with additive noise ϵ_k ,

$$\mathbf{y}_k = f_k(\mathbf{x}_k) + \epsilon_k \quad (2)$$

Following a standard Bayesian approach [18] priors are placed over the latent functions and variables. Where the prior encodes our knowledge of the generating process *before* the model is informed by data; then, by conditioning (or updating) the priors given measurements, we find the posterior distribution. The posterior can then be used to make predictions, which utilise a combination of (i) domain knowledge encoded in the prior and (ii) patterns in the observed data.

First, a shared Gaussian process (GP) prior distribution is placed over each latent function,

$$f_k \sim \text{GP}(m(\cdot), k(\cdot, \cdot)) \quad (3)$$

This prior is fully specified by its mean $m(\cdot)$ and covariance functions $k(\cdot, \cdot)$. The mean captures the expected *path* of the response, while the covariance describes the similarity between any two outputs, with respect to their input (i.e. the function's *smoothness*). Domain expertise and knowledge of the underlying physics can (and should) be encoded

within both of these functions; although, for now, we keep their definitions simple,

$$m(x_i) = 0 \quad (4)$$

$$k(x_i, x_j) = \alpha^2 \left(1 + \frac{\sqrt{3} |\mathbf{x}_i - \mathbf{x}_j|}{l} \right) \exp \left(-\frac{\sqrt{3} |\mathbf{x}_i - \mathbf{x}_j|}{l} \right) \quad (5)$$

these correspond to a zero-mean and Matérn 3/2 kernel function. These prior functions have *hyperparameters*: the *process variance* α and *length scale*, l which have their own (hyper) priors,

$$\alpha \sim \text{Truncated-Normal}_+(1, 1), \quad l \sim \text{Half-Normal}_+(1) \quad (6)$$

The prior distribution for the additive noise is also defined, and collected with other latent variables in $\boldsymbol{\theta}$,

$$\epsilon_{ki} \sim \text{Normal}(0, \sigma_k), \quad \sigma_k \sim \text{Half-Normal}_+ (.2) \quad (7)$$

$$\boldsymbol{\theta} \triangleq \{\alpha, l, \{\sigma_k\}_{k=1}^K\} \quad (8)$$

Note, the half-normal distributions are zero-centred and the truncated normal has a minimum value of zero, since $\{\alpha, l, \{\sigma_k\}\}$ must be non-negative.

Since the hyperparameters also have a Bayesian treatment, they are also updated given the training data. Practically, this means that the shape of the prior adapts during inference (via the mean and covariance functions). The above priors are set to be weakly informative, in view of the normalised space, postulating a signal-to-noise ratio of 5 (in terms of standard deviation).

Importantly, k subscripts are missing for $\{\alpha, l\}$, indicating that hyperparameters are *shared* between all k GPs (of both 350 and 22x events). A convenient interpretation is that the shared prior captures the characteristics of *typical* commuter trains, which should be useful in monitoring procedures (especially novelty detection). Looking forward, $m(\cdot)$ should be specified beyond a zero-mean assumption (with parameters contributing to $\boldsymbol{\theta}$) to present a general mean function, which might represent a generic 350 train or 22x train specifically. The semi-parametric formulation Kennedy and O’Hagan [19] will be considered, alongside a basis function interpretation [17].

A Low-rank Implementation to Encode Boundary Conditions

The number of observations per event varies quite widely, depending on the speed of the train (750-1500). Since Gaussian processes scale badly with large data, we implement a low-rank approximation. Following Riutort-Mayol *et al.* [20] another basis function approximation is used for the *kernel function*, which reduces the order of complexity from $O(N^3)$ to $O(N \cdot M + M)$, where M is the number of basis functions.

As the Matérn 3/2 kernel is a stationary function, it can be approximated using an eigen-decomposition. Solin and Särkkä [21] achieve this by interpreting $k(\cdot, \cdot)$ as the kernel of a pseudo-differential operator. The pseudo-differential operator is then approximated via

Hilbert space methods on a compact subset $\Omega \subset \mathbb{R}^D$, subjected to boundary conditions. A convenient side-effect is that the boundary conditions can be utilised to constrain the approximation given domain knowledge, see Jones *et al.* [22] for an engineering example.

Considering the z-score normalised inputs in Figure 5, we know that most inputs should vary between approximately $[-2, 2]$ while returning to zero-strain at the boundaries of the envelope. This scaling naturally defines an interval in which the approximated GP is valid,

$$\Omega \in [-L, L] = [-3, 3] \quad (9)$$

Note the *buffer* for L (such that $L > 2$) is to support stable inferences and ensure that test data do not fall outside the approximating window of Ω .

Within Ω , the kernel is then be approximated as follows [20],

$$k(x_i, x_j) \approx \sum_{m=1}^M S(\sqrt{\lambda_m}) \phi_m(x_i) \phi_m(x_j) \quad (10)$$

where $x_i, x_j \in \Omega$; S is the spectral density of the kernel function; λ_m are the eigenvalues and $\phi_m(x)$ are the eigenvectors of the Laplacian operator in the given domain Ω . In these experiments, we approximate with 40 basis functions ($M = 40$).

For a univariate Matérn 3/2 function, $S(\cdot)$ is as follows [21],

$$S(\omega) = \alpha \frac{4 \cdot 3^{3/2}}{l^3} \left(\frac{3}{l^2} + \omega^\top \omega \right)^{-2} \quad (11)$$

Then, following [20], the following eigenvalue problem is considered, with *Dirichlet* boundary conditions respecting Ω ,

$$-\nabla^2 \phi_m(x) = \lambda_m \phi_m(x), \quad x \in \Omega \quad (12)$$

$$\phi_m(x) = 0, \quad x \notin \Omega \quad (13)$$

The solution of which presents the following analytical eigenvalues and eigenvectors,

$$\lambda_m = \left(\frac{m\pi}{2L} \right)^2 \quad (14)$$

$$\phi_m(x) = \sqrt{\frac{1}{L}} \sin \left(\sqrt{\lambda_m} (x + L) \right) \quad (15)$$

Practically, Dirichlet conditions are justified, since the strain envelopes are bounded to zero before and after each signal (in view of the normalisation and strain-conversion). Further details of the equivalent linear implementation can be found in [20].

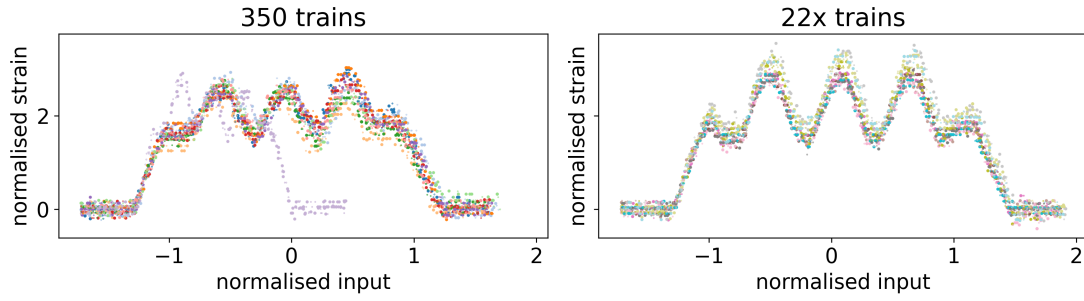


Figure 5. Ten 350-type and ten 22x-type train passing events: all modelled by the multilevel GP regression. Events are plotted with different colours. Pink markers in the left-hand plot correspond to an outlying event.

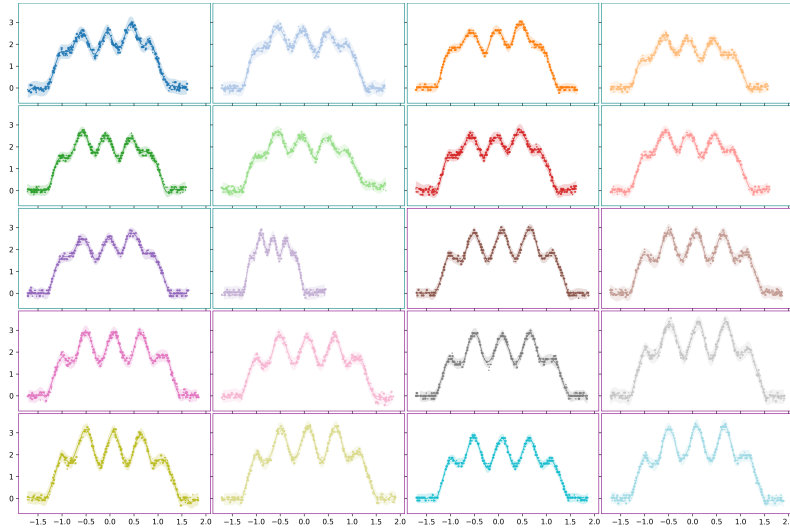


Figure 6. Posterior predictive distributions for the multilevel GP: teal frames are for the 350 train types and purple frames are for 22x trains.

RESULTS: TOWARDS MONITORING PROCEDURES

The dataset comprises 20 passing events: ten 350 trains and ten 22x trains, sampled at random from the historical data and plotted in Figure 5. We intentionally include an outlying event, shown by the pink markers in the left panel. These outlying data appear to relate to an erroneous speed measurement, and therefore, an incorrect scaling of the inputs. This outlier (within the 350 trains) is used to demonstrate simple monitoring capabilities.

The models were written in probabilistic programming language *Stan* [23] and inferred via MCMC with the no U-turn implementation of Hamiltonian Monte Carlo [24]. Throughout, 1000 iterations are used for burn-in and inference. The posterior predictive distribution of each GP is shown in Figure 6. The plots indicate a good fit to the data, however, the representation at the start/end of each strain envelope is quite poor, since the GP must represent the entire function without a parametrised mean – as discussed, this will be the focus of future work.

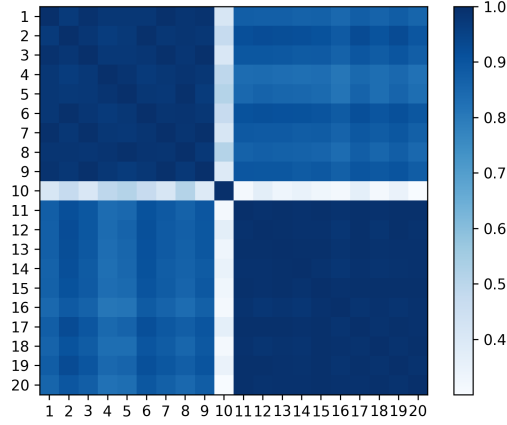


Figure 7. Pairwise Pearson’s correlation between the (expected) posterior distribution of each event GP in the combined model.

Rather than focus on prediction metrics here, we present how the correlation structure might inform monitoring procedures. One simple approach plots the correlation between the (expectation) of each f_k posterior in the combined model, presented in Figure 7. Intuitively, there are two blocks within this confusion matrix: one associated with each train type. Additionally, the abnormal 350-event (index 9 within the upper left block) would clearly be identified as outlying: data which are not flagged by the decision tree heuristic (Figure 3).

CONCLUDING REMARKS

We have presented initial work towards modelling multilevel data structures, actively streaming from the (IB5) Staffordshire rail bridge. Firstly, a pre-existing bridge-weight-in-motion system was used to classify train events into groups of a specific vehicle type – here, we focus on commuter trains. For repeated crossings, the data were treated as longitudinal (or panel) data. A hierarchical (low-rank approximation) Gaussian process was then used to model aggregate train events, working towards a general representation of the strain response.

Looking forward, it would be useful to parametrise the mean function, to learn a *typical* representation for each *train type*, alongside event-specific models. A basis function formulation using *strain influence lines* will be investigated. An interpretable mean function should enable further insights regarding the variation of strain response, as well as the option for more expressive simulations.

ACKNOWLEDGEMENTS

The authors would like to thank Professor C.R. Middleton, Dr. F. Huseynov and P.R.A. Fiddler from the University of Cambridge, who led the latest phase of the instrumentation of the Norton Bridge, thanks to the support of the Centre for Digital Built Britain’s

(CDBB). CDBB is a core partner of the Construction Innovation Hub, funded by UK Research and Innovation (UKRI) through the Industrial Strategy Challenge Fund (ISCF).

LAB and MG acknowledge the support of the UK Engineering and Physical Sciences Research Council (EPSRC) through ROSEHIPS (EP/W005816/1). AD is supported by Wave 1 of The UKRI Strategic Priorities Fund under the EPSRC Grant EP/T001569/1 and EPSRC Grant EP/W006022/1, particularly the *Ecosystems of Digital Twins* theme within those grants & The Alan Turing Institute.

Finally, this research was possible thanks to the support of the 2022 Trimble Fund of The University of Cambridge.

REFERENCES

1. L. J. Butler, N. Gibbons, C. Middleton, and M. Z. Elshafie. Integrated fibre-optic sensor networks as tools for monitoring strain development in bridges during construction. In *International Association of Bridge and Structural Engineering*, 2016.
2. L. J. Butler, W. Lin, J. Xu, N. Gibbons, M. Z. Elshafie, and C. R. Middleton. Monitoring, modelling, and assessment of a self-sensing railway bridge during construction. *Journal of Bridge Engineering*, 23(10):04018076, 2018.
3. E. Cross. *On structural health monitoring in changing environmental and operational conditions*. PhD thesis, University of Sheffield, 2012.
4. N. Dervilis, K. Worden, and E. Cross. On robust regression analysis as a means of exploring environmental and operational conditions for shm data. *Journal of Sound and Vibration*, 347:279–296, 2015.
5. H. Sohn, C. R. Farrar, F. M. Hemez, D. D. Shunk, D. W. Stinemates, B. R. Nadler, and J. J. Czarnecki. A review of structural health monitoring literature: 1996–2001. *Los Alamos National Laboratory, USA*, 1:16, 2003.
6. L. Bull, P. Gardner, J. Gosliga, T. Rogers, N. Dervilis, E. Cross, E. Papatheou, A. Maguire, C. Campos, and K. Worden. Foundations of population-based SHM, part I: Homogeneous populations and forms. *Mechanical Systems and Signal Processing*, 148:107141, 2021.
7. J. Gosliga, P. Gardner, L. Bull, N. Dervilis, and K. Worden. Foundations of population-based SHM, part II: Heterogeneous populations—graphs, networks, and communities. *Mechanical Systems and Signal Processing*, 148:107144, 2021.
8. P. Gardner, L. Bull, J. Gosliga, N. Dervilis, and K. Worden. Foundations of population-based SHM, part III: Heterogeneous populations—mapping and transfer. *Mechanical Systems and Signal Processing*, 149:107142, 2021.
9. L. A. Bull, P. Gardner, T. J. Rogers, N. Dervilis, E. J. Cross, E. Papatheou, A. E. Maguire, C. Campos, and K. Worden. Bayesian modelling of multivalued power curves from an operational wind farm. *Mechanical Systems and Signal Processing*, 169:108530, 2022.
10. G. P. Tsialiamanis, D. Wagg, P. A. Gardner, N. Dervilis, and K. Worden. On partitioning of an shm problem and parallels with transfer learning. In *Topics in Modal Analysis & Testing, Volume 8: Proceedings of the 38th IMAC, A Conference and Exposition on Structural Dynamics 2020*, page 41. Springer Nature, 2020.

11. L. Bull, D. Di Francesco, M. Dhada, O. Steinert, T. Lindgren, A. K. Parlikad, A. Duncan, and M. Girolami. Hierarchical bayesian modelling for knowledge transfer across engineering fleets via multitask learning. *Computer-Aided Civil and Infrastructure Engineering*, 38(7): 821–848, 2023.
12. A. Anaissi, B. Suleiman, and W. Alyassine. Personalised federated learning framework for damage detection in structural health monitoring. *Journal of Civil Structural Health Monitoring*, 13(2-3):295–308, 2023.
13. W. Lin, L. J. Butler, M. Z. Elshafie, and C. R. Middleton. Performance assessment of a newly constructed skewed half-through railway bridge using integrated sensing. *Journal of Bridge Engineering*, 24(1):04018107, 2019.
14. D. G. Broo, M. Bravo-Haro, and J. Schooling. Design and implementation of a smart infrastructure digital twin. *Automation in Construction*, 136:104171, 2022.
15. P. Fidler, F. Huseynov, M. Bravo Haro, V. Vilde, J. Schooling, and C. Middleton. Augmenting an existing railway bridge monitoring system with additional sensors to create a bridge weigh-in-motion system and digital twin. In *SHMII-11: 11th International Conference on Structural Health Monitoring of Intelligent Infrastructure*, 2022.
16. F. Huseynov, P. Fidler, M. Bravo Haro, V. Vilde, J. Schooling, and C. Middleton. Setting up a real-time train load monitoring system in the uk using bridge weigh-in motion technology-a case study. In *SHMII-11: 11th International Conference on Structural Health Monitoring of Intelligent Infrastructure*, 2023.
17. E. J. OBrien, M. Quilligan, and R. Karoumi. Calculating an influence line from direct measurements. In *Proceedings of the Institution of Civil Engineers-Bridge Engineering*, volume 159, pages 31–34. Thomas Telford Ltd, 2006.
18. C. E. Rasmussen, C. K. Williams, *et al.* *Gaussian Processes for Machine Learning*, volume 1. Springer, 2006.
19. M. C. Kennedy and A. O’Hagan. Bayesian calibration of computer models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(3):425–464, 2001.
20. G. Riutort-Mayol, P.-C. Bürkner, M. R. Andersen, A. Solin, and A. Vehtari. Practical hilbert space approximate bayesian gaussian processes for probabilistic programming. *Statistics and Computing*, 33(1):17, 2023.
21. A. Solin and S. Särkkä. Hilbert space methods for reduced-rank gaussian process regression. *Statistics and Computing*, 30(2):419–446, 2020.
22. M. R. Jones, T. J. Rogers, and E. J. Cross. Constraining gaussian processes for physics-informed acoustic emission mapping. *Mechanical Systems and Signal Processing*, 188: 109984, 2023.
23. B. Carpenter, A. Gelman, M. D. Hoffman, D. Lee, B. Goodrich, M. Betancourt, M. Brubaker, J. Guo, P. Li, and A. Riddell. Stan: A probabilistic programming language. *Journal of statistical software*, 76(1), 2017.
24. M. D. Hoffman, A. Gelman, *et al.* The no-u-turn sampler: adaptively setting path lengths in hamiltonian monte carlo. *J. Mach. Learn. Res.*, 15(1):1593–1623, 2014.