Towards practical Bayesian system identification of engineering structures with spatially dense measurements

Martínez, Andrés; Slobbe, Arthur; Rózsás, Árpád; Rocha, Iuri; van der Meer, Frans

DOI
10.1016/j.engstruct.2025.120214

Publication date
2025

Document Version
Final published version

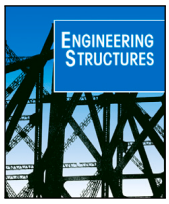
Published in
Engineering Structures

Engineering Structures 334 (2025) 120214

Contents lists available at ScienceDirect

Engineering Structures

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



Towards practical Bayesian system identification of engineering structures with spatially dense measurements

Andrés Martínez [a,*], Arthur Slobbe [b], Árpád Rózsás [a], Iuri Rocha [c], Frans van der Meer [c]

[a] Department of Reliable Structures, TNO, Delft, Netherlands
[b] Department of Building Materials & Structures, TNO, Delft, Netherlands
[c] Faculty of Civil Engineering and Geosciences, Delft University of Technology, Delft, Netherlands

ARTICLE INFO

Keywords:
Structural Health Monitoring
Bayesian system identification
Variational Bayesian Monte Carlo
Distributed fibre optic sensing
Spatial correlation
Bayesian model selection

ABSTRACT

Bayesian system identification is increasingly used in Structural Health Monitoring (SHM) to infer unobservable parameters of a structure from sensor data. The use of spatially dense measurements, such as those from distributed fibre optic sensors, can further enhance the results of Bayesian system identification due to the large volume of data. However, this combination faces two major challenges: the computational cost of inference and the correlation structure of closely spaced data points. To overcome these difficulties, we propose a methodology that combines the recently-developed Variational Bayes Monte Carlo (VBMC) method with Gaussian process modelling of model discrepancy, and extend VBMC to enable posterior predictive calculations without additional model evaluations. We demonstrate the effectiveness of the proposed methodology on a reinforced concrete slab bridge instrumented with distributed fibre optic strain sensors and analysed using a finite element model. The main outcome is that VBMC requires fewer than 200 finite element model evaluations while producing accurate estimates, whereas a conventional MCMC method requires thousands. The application of the proposed framework provides two additional novel insights: accounting for spatial correlations improves model performance and higher measurement resolution leads to more precise parameter estimates, though with limited impact on predictive accuracy. This study advances the practical implementation of Bayesian system identification in SHM by providing both the computational efficiency and statistical framework needed for modern sensing technologies.

1. Introduction

1.1. Motivation

Government agencies around the world are facing enormous renovation and replacement tasks since a significant portion of their infrastructure was built decades ago and is reaching the end of its service life. Traditional methods of assessing structural condition, such as manual inspections, are not only costly but also rely on specialized labour. As an alternative, Structural Health Monitoring (SHM) has emerged as a more efficient and consistent approach to make informed decisions for infrastructure maintenance.

The use of high-resolution sensors in SHM has gained considerable attention in recent years. Technologies such as distributed fibre optic sensors (DFOS) [1], digital image correlation (DIC) [2] and Micro-Electro-Mechanical Systems (MEMS) sheets [3] enable the collection of data with unprecedented spatial and/or temporal detail. These sensors provide a wealth of information that has the potential to improve the overall effectiveness and accuracy of SHM.

Occasionally, engineers use sensor data to calibrate structural models, treating SHM as an inverse problem where the objective is to estimate unobserved parameters of a physical model from measurements. Inverse problems are usually challenging due to the presence of observational noise, model discrepancies and the non-uniqueness of solutions [4]. Bayesian system identification offers an effective probabilistic framework for addressing inverse problems by providing several advantages over traditional optimization methods: it allows for the incorporation of prior knowledge about the parameters, it inherently regularizes ill-posed problems, and it quantifies uncertainties in the parameter estimates. Under this approach, the unknown structural parameters are treated as random variables with a prior distribution that is updated to a posterior using the likelihood of the observed data.

To fully exploit the data from high spatial resolution sensors, one might employ Bayesian system identification. However, this combination presents significant challenges. First, computationally expensive high-fidelity models are usually required to provide information at

* Corresponding author.
E-mail address: andres.martinezcolan@tno.nl (A. Martínez).

https://doi.org/10.1016/j.engstruct.2025.120214
Received 21 November 2024; Received in revised form 8 February 2025; Accepted 25 March 2025
Available online 5 April 2025

the same level of detail as the sensor data, making common Bayesian methods like Markov Chain Monte Carlo (MCMC) prohibitive, as they require numerous model evaluations. Second, model discrepancies—differences between simulation and reality—can exhibit strong correlations in closely spaced data points, but the optimal approach for modelling this correlation structure remains uncertain. While neglecting these correlations can result in flawed inferences, explicitly modelling them adds complexity to the problem, as it introduces additional statistical parameters that must be inferred. In optimization-based approaches, this typically results in a highly nonlinear objective function, making the problem challenging. Bayesian inference naturally accommodates this complexity by implicitly regularizing the problem through the prior distribution. Addressing these computational and statistical challenges is essential to enable the practical application of Bayesian system identification of engineering structures with modern sensing technologies.

1.2. Related work

One common approach to overcome the computational barrier is to replace structural model responses with cheap-to-compute surrogate models. They are built by strategically selecting samples and fitting an approximation function with input–output pairs. Various types of surrogates have been used for system identification of engineering structures such as response surfaces [5,6], Gaussian processes [7–10], polynomial chaos expansions [11–13] and artificial neural networks [14–16]. A key challenge with surrogate models is the amount of simulations required to obtain the desired accuracy, which can become unmanageable. This is especially true for high-dimensional input spaces, as many surrogate methods are prone to the curse of dimensionality [17].

Other methods achieve computational efficiency through specialized formulations. For instance, Titscher et al. [18] combines mean field variational inference with first-order Taylor expansions, while Febrianto et al. [19] employs the statFEM methodology with simplified stochastic partial differential equations. Although these approaches require few model evaluations, their applicability is limited by their underlying assumptions. Indeed, the former struggles with nonlinear physical models, and the latter relies on specific mathematical simplifications that may not generalize well to other structural systems

Recent advancements in Bayesian inference have focused on directly approximating the unnormalized posterior to accelerate the process. By concentrating on regions that significantly influence the posterior distribution, these techniques greatly reduce the number of required evaluations. Methods such as Bayesian Active Posterior Estimation (BAPE) [20] and Adaptive Gaussian Process (AGP) [21] employ Gaussian process models to surrogate the unnormalized posterior. The Variational Bayesian Monte Carlo (VBMC) method [22,23] adopts a similar strategy, but with a variational framework that approximates the posterior with a parametric distribution.

Moving to spatial correlations, their treatment in Bayesian system identification was first introduced in the seminal work of Kennedy and O'Hagan [24], who proposed the use of Gaussian processes for model discrepancy, with the dependency being encoded in its correlation function. Brynjarsdóttir and O'Hagan [25] used the same approach and demonstrated that only when incorporating realistic priors into the model discrepancy function, the true parameter values are uncovered. Recent developments have focused on specialized formulations. Ramancha et al. [13] introduced a covariance function tailored for linear dynamic systems based on the theory of random vibrations. Kosikova et al. [26] studied the automatic selection of a covariance function with an efficient approach based on Laplace approximations. Koune et al. [27] proposed an efficient likelihood evaluation framework by exploiting separable spatio-temporal covariance structures and utilizing exponential functions. Although effective, these approaches rely on specific constraints and assumptions that might limit their broader applicability across different types of engineering structures.

1.3. Present study and contributions

A key gap in the SHM literature is the lack of efficient Bayesian methods for real-world structures with spatially dense measurements. While efficient inference techniques exist, they often rely on restrictive assumptions about the model structure or correlation patterns that limit their practical applicability.

To address this gap, we propose the use of the VBMC method, as it offers several advantages that make it particularly suitable for this application: it requires minimal likelihood evaluations, making it practical for computationally expensive models; it imposes no restrictions on model structure or correlation patterns; it provides approximations for both the posterior distribution and model evidence, enabling Bayesian model selection; and it has a readily available Python implementation (PyVBMC) [28], which requires no custom coding or extensive fine-tuning. Additionally, we extend VBMC by presenting a practical approach for computing posterior predictive distributions without requiring additional model evaluations.

Our methodology combines VBMC with a flexible treatment of model discrepancy. Similarly to Kennedy and O'Hagan, we represent discrepancy as a Gaussian process with a distance-based covariance function; however, we perform full Bayesian inference on both physical and correlation parameters, which allow us to learn the correlation structure from the data. Furthermore, unlike previous approaches, we maintain generality by avoiding restrictions on the covariance function.

We demonstrate our methodology on a reinforced concrete bridge using static strain measurements from DFOS. To our knowledge, this represents the first application of VBMC for inference in a real-world structure. Additionally, we use this framework to investigate two aspects that have received limited attention in the literature. First, we evaluate different covariance models to investigate the effect of ignoring correlations on inference, and to provide insights for choosing a specific function. Second, we examine the value of high-resolution measurements for inference, as the relationship between measurement density and parameter uncertainty becomes non-trivial in the presence of spatial dependencies.

1.4. Organization

The paper is structured as follows: Section 2 details the methodology, including the VBMC method and the modelling of spatial correlations. Section 3 describes the case study of the reinforced concrete bridge. Sections 4 to 6 present our findings, focusing on performance assessment, covariance function evaluation, and the impact of data resolution, respectively. Sections 7 and 8 contain the discussion and conclusions.

2. Methodology

2.1. Principles of Bayesian inference

Bayesian inference provides a formal framework for updating the probability distributions of uncertain parameters given observations through the application of Bayes' theorem. For continuous variables, this theorem takes the form:

$$p(\boldsymbol{\theta}|\boldsymbol{y}) = \frac{p(\boldsymbol{y}|\boldsymbol{\theta}) \cdot p(\boldsymbol{\theta})}{\int p(\boldsymbol{y}|\boldsymbol{\theta}) \cdot p(\boldsymbol{\theta})\, d\boldsymbol{\theta}} \tag{1}$$

The posterior distribution $p(\boldsymbol{\theta}|\boldsymbol{y})$ represents our updated knowledge about the parameters $\boldsymbol{\theta}$ after observing data $\boldsymbol{y}$. This combines our prior knowledge, encoded in $p(\boldsymbol{\theta})$, with the information from the data through the likelihood function $p(\boldsymbol{y}|\boldsymbol{\theta})$. The denominator, known as the evidence or marginal likelihood, ensures the posterior distribution integrates to one over the parameter domain.

For most practical applications, the integral in the denominator is intractable and numerical approaches are needed. Markov Chain

Monte Carlo (MCMC) methods have become the standard in Bayesian statistics as they can generate samples from the posterior distribution without requiring the explicit computation of the evidence. However, the computational cost of generating a sufficient number of samples for convergence, typically thousands, might be prohibitively high in cases with expensive likelihood function evaluations. In the next section, we introduce Variational Bayesian Monte Carlo as an efficient alternative to MCMC for engineering structures.

2.2. Variational Bayesian Monte Carlo

Variational inference methods provide an efficient alternative to sampling-based methods by framing Bayesian inference as an optimization problem. They approximate the true posterior distribution with a simpler, parametric distribution q. The parameters ϕ of this distribution are optimized to maximize the Evidence Lower Bound (ELBO), which is equivalent to minimizing the Kullback–Leibler divergence between the approximate and true posterior distributions. The ELBO is defined as follows:

$$\text{ELBO} = \mathbb{E}_{q_\phi}[\log p(\boldsymbol{y}, \boldsymbol{\theta})] + \mathcal{H}[q_\phi(x)] \tag{2}$$

where the first term is the expected log joint probability for q_ϕ, and the second term is the entropy of the variational posterior.

Building on this framework, Variational Bayesian Monte Carlo (VBMC) [22,23] combines variational inference with Gaussian Process surrogate modelling to efficiently approximate posterior distributions in scenarios where likelihood evaluations are computationally expensive. The method provides both an approximation to the posterior distribution and an estimate of the model evidence through the ELBO. In VBMC, the distribution q is a mixture of Gaussians:

$$q_\phi(\boldsymbol{\theta}) = \sum_{k=1}^{K} w_k \mathcal{N}(\boldsymbol{\theta}; \boldsymbol{\mu}_k, \sigma_k^2 \boldsymbol{\Sigma}) \tag{3}$$

New samples are selected by maximizing this acquisition function:

$$a(\boldsymbol{\theta}) = V_{GP}(\boldsymbol{\theta}) q_\phi(\boldsymbol{\theta}) \exp(m_{GP}(\boldsymbol{\theta})) \tag{4}$$

where $V_{GP}(\boldsymbol{\theta})$ is the posterior predictive variance of the GP, $m_{GP}(\boldsymbol{\theta})$ is its posterior predictive mean, and $q_\phi(\boldsymbol{\theta})$ refers to the current variational posterior. This function promotes sampling in regions of high GP uncertainty (V_{GP}), areas of high posterior probability (q_ϕ), and locations that the GP predicts might be promising ($\exp(m_{GP})$).

The variational approximation is refined through an iterative process that follows these steps:

1. Add samples at promising locations in the parameter space by maximizing the acquisition function
2. Train a Gaussian Process surrogate of the log joint probability (sum of log prior and log likelihood) using all collected samples
3. Update the variational parameters by optimizing the ELBO, where the expectation term is computed in closed form using the GP approximation, and the entropy is estimated via Monte Carlo sampling

Initially, VBMC uses a warm-up phase with $K = 2$ mixture components and equal weights. This phase continues until the improvement in the ELBO across consecutive iterations falls below a threshold for several iterations, indicating initial stability of the approximation. After warm-up, VBMC begins adjusting the complexity of the variational posterior by adding new components through splitting existing ones when the ELBO shows consistent improvement, and removing components with negligible weights if their removal does not significantly impact the approximation quality. This process continues until convergence in the ELBO. For a detailed description of the method, the reader is referred to the original papers by Acerbi [22,23].

2.3. Data generating process with spatial correlations

To define the likelihood function, one must construct a narrative that describes the data generating process [29]. In this paper, the data generating process is conceptualized as the combination of a physical model and a probabilistic model. The physical model is a deterministic numerical simulation of a structure for which responses can be obtained as a function of control variables, such as location, and physical parameters; while the probabilistic model accounts for the intrinsic uncertainties that appear when using a model to describe a physical system. Here, we consider two sources of uncertainty: measurement uncertainty and physical model uncertainty.

Measurement uncertainty refers to the mismatch between the measured quantities and the true responses of the structure caused by sensing errors and/or environmental noise. In contrast, model uncertainty relates to the discrepancy between the deterministic physical model predictions and the true responses, arising from mathematical simplifications and idealizations in the physical model. We adopt the model introduced by Kennedy and O'Hagan [24]:

$$y_i = \zeta(x_i) + \epsilon_i \tag{5}$$

where y_i is the ith observation of the physical system, $\zeta(x_i)$ is the true value of the physical system at location x_i and ϵ_i is the independent measurement error. The model discrepancy is represented as an additive term $\delta(x)$ on the numerical simulator $\eta(x, \boldsymbol{\theta}_p)$, with $\boldsymbol{\theta}_p$ the unknown physical parameter vector:

$$\zeta(x) = \eta(x, \boldsymbol{\theta}_p) + \delta(x) \tag{6}$$

Combining Eqs. (5) and (6) yields the data generating process equation:

$$y_i = \eta(x_i, \boldsymbol{\theta}_p) + \delta(x_i) + \epsilon_i, \quad i = 1, \ldots, n \tag{7}$$

The measurement error terms ϵ_i are modelled as independent zero-mean Gaussian variables with a standard deviation σ_ϵ, which can either be directly inferred or estimated beforehand, for instance, using sensor pre-recordings. On the other hand, the model discrepancy function is represented as a zero-mean Gaussian process:

$$\delta(x) \sim GP(\boldsymbol{0}, k(x, x')) \tag{8}$$

where $k(x, x')$ is a covariance function. Since we are interested in cases with a high spatial density of points, it is logical to employ a distance-based covariance function, which exhibits strong correlations for nearby points and weaker correlations for those far apart. Furthermore, since the regions of higher or lower model discrepancy are not always known in advance, a stationary covariance function is a prudent choice. The Matérn covariance function is a commonly used example of a distance-based stationary function. It is characterized by the parameters σ_a (the scale term), l (the correlation length), and ν, which controls the smoothness of the process. For $\nu = 1.5$, the kernel function reads:

$$k(x, x') = \sigma_a^2 \left(1 + \frac{\sqrt{3}|x - x'|}{l}\right) \exp\left(-\frac{\sqrt{3}|x - x'|}{l}\right) \tag{9}$$

A fundamental property of Gaussian processes is that any finite collection of random variables sampled from the process follows a multivariate Gaussian distribution [30]. Therefore, the measurement vector follows:

$$\boldsymbol{y} \sim \mathcal{N}(\eta(\boldsymbol{x}, \boldsymbol{\theta}_p), \boldsymbol{K} + \sigma_\epsilon^2 \boldsymbol{I}) \tag{10}$$

where $\boldsymbol{K}$ is the covariance matrix of the model discrepancy function $\delta(x)$. The elements of matrix $\boldsymbol{K}$ are defined using the covariance function $k(x, x')$:

$$K_{ij} = k(x_i, x_j) \tag{11}$$

Defining $\boldsymbol{\Sigma} = \boldsymbol{K} + \sigma_\epsilon^2 \boldsymbol{I}$ as the total covariance matrix and n as the number of observations, the likelihood function is the density of the multivariate Gaussian distribution:

$$p(\boldsymbol{y}|\boldsymbol{\theta}) = \frac{1}{(2\pi)^{n/2} |\boldsymbol{\Sigma}|^{1/2}} \exp\left(-\frac{1}{2}(\boldsymbol{y} - \eta(\boldsymbol{x}, \boldsymbol{\theta}_p))^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{y} - \eta(\boldsymbol{x}, \boldsymbol{\theta}_p))\right) \tag{12}$$

Note that $\boldsymbol{\theta}$ includes the physical parameters $\boldsymbol{\theta}_p$, which govern the physical model η, as well as the statistical parameters $\boldsymbol{\theta}_s$, which determine the total covariance matrix $\boldsymbol{\Sigma}$.

2.4. Post-inference applications

Although parameter inference might sometimes be a goal in itself, in most occasions we would like to use the posterior distribution for further analysis. In this section, we explore three applications and show how to adapt the results from VBMC for these purposes.

2.4.1. Posterior predictive distribution

Besides estimating the underlying parameters of our model, we often want to make predictions based on these parameters. If we knew the exact value of $\boldsymbol{\theta}$, predicting would be straightforward. However, since we only have the posterior distribution of possible values for $\boldsymbol{\theta}$, we need to marginalize over this distribution to make predictions [31].

$$p(\tilde{\boldsymbol{y}}|\boldsymbol{y}) = \int p(\tilde{\boldsymbol{y}}|\boldsymbol{\theta}) \cdot p(\boldsymbol{\theta}|\boldsymbol{y})\, d\boldsymbol{\theta} \tag{13}$$

The term $p(\tilde{\boldsymbol{y}}|\boldsymbol{y})$ is known as the posterior predictive distribution of $\tilde{\boldsymbol{y}}$ given $\boldsymbol{y}$, and $p(\tilde{\boldsymbol{y}}|\boldsymbol{\theta})$ is the likelihood of the unobserved data. The posterior predictive is the expected value of the likelihood of the unobserved data over the posterior; hence, it incorporates the uncertainty from all sources: model and measurement uncertainty via the likelihood, and parameter uncertainty via the posterior.

In practice, the integral in Eq. (13) is often intractable, but it can be approximated in a Monte Carlo way. Let $\boldsymbol{\theta}^{(i)}$ represent samples from the posterior distribution $p(\boldsymbol{\theta}|\boldsymbol{y})$:

$$p(\tilde{\boldsymbol{y}}|\boldsymbol{y}) \approx \frac{1}{N} \sum_{i=1}^{N} p(\tilde{\boldsymbol{y}}|\boldsymbol{\theta}^{(i)}) \tag{14}$$

Instead of evaluating the posterior predictive distribution for every conceivable value of $\tilde{\boldsymbol{y}}$, we opt for directly generating samples from the conditional distribution of the unobserved data given each posterior sample $\boldsymbol{\theta}^{(i)}$. Let $\tilde{\boldsymbol{y}}^{(i)}$ denote a sample drawn from this conditional distribution:

$$\tilde{\boldsymbol{y}}^{(i)} \sim p(\tilde{\boldsymbol{y}}|\boldsymbol{\theta}^{(i)}) = \mathcal{N}(\eta(\tilde{\boldsymbol{x}}, \boldsymbol{\theta}_p^{(i)}), \tilde{\boldsymbol{K}} + \sigma_\epsilon^2 \boldsymbol{I}) \tag{15}$$

The collection of these samples $\{\tilde{\boldsymbol{y}}^{(i)}\}_{i=1}^{N}$ forms an approximation to the posterior predictive distribution. However, this approach can be extremely inefficient, as each sample implies evaluating the physical model $\eta(\tilde{\boldsymbol{x}}, \boldsymbol{\theta}_p^{(i)})$, demanding numerous additional simulations to propagate the uncertainty of the posterior distribution to predictions. This limitation can restrict the applicability of VBMC.

To address this problem, we propose an extension of VBMC for estimating the posterior predictive distribution without additional model evaluations. Suppose that for each point $\boldsymbol{\theta}^{(j)}$ where VBMC evaluates the true likelihood we calculate the model response not only at the observed locations $\boldsymbol{x}$ but also at the prediction locations $\tilde{\boldsymbol{x}}$. By storing these additional responses, we can train a surrogate model. Because this surrogate is built with the samples VBMC uses to represent the posterior, it is expected to be accurate in regions with high posterior density, which is essential for obtaining a good approximation of the posterior predictive distribution. Here, we use a zero-mean Gaussian process surrogate and tune its hyperparameters by marginal likelihood optimization. For the case study, we use only the mean predictions from the GP as the predicted variances were negligible. If this were not the case, Eqs. (13) to (15) can be extended by adding a surrogate uncertainty term, using the predicted GP standard deviation.

It should be noted that the approach described in this section is heuristic, and there is no formal guarantee of posterior predictive accuracy; however, it delivered excellent results in our experiments.

2.4.2. Bayesian model selection

When multiple competing models can explain the observed data, Bayesian model selection provides a principled framework for comparison. In this work, we compare different correlation models for the discrepancy function from Eq. (8). The Bayes factor quantifies the relative support for one model over another:

$$B_{12} = \frac{p(\boldsymbol{y}|M_1)}{p(\boldsymbol{y}|M_2)} = \frac{\int p(\boldsymbol{y}|\boldsymbol{\theta}_1, M_1) \cdot p(\boldsymbol{\theta}_1|M_1)\, d\boldsymbol{\theta}_1}{\int p(\boldsymbol{y}|\boldsymbol{\theta}_2, M_2) \cdot p(\boldsymbol{\theta}_2|M_2)\, d\boldsymbol{\theta}_2} \tag{16}$$

where M_1 and M_2 represent the competing models. An advantage of using Bayes factors is that they penalize model complexity, as models with more parameters require the likelihood to be integrated over a larger parameter space. In this work, we follow Jeffreys' interpretation [32] shown in Table 1 to qualify the strength of evidence of one model over the other.

Table 1
Strength of evidence of M_1 over M_2 based on the Bayes factor [32].

B_{12}	Strength of evidence
$<10^0$	Negative
10^0 to $10^{1/2}$	Barely worth mentioning
$10^{1/2}$ to 10^1	Substantial
10^1 to $10^{3/2}$	Strong
$10^{3/2}$ to 10^2	Very strong
$>10^2$	Decisive

In the context of VBMC, we use the evidence lower bound (ELBO) as a proxy for the model evidence. This can be justified because the tightness of this bound is explicitly optimized during the variational inference process. The Bayes factor is then computed as the ratio of the ELBOs obtained during inference under the different models.

2.4.3. Information content of observations

The Kullback–Leibler (KL) divergence provides a natural way to quantify the information content of observations by measuring the *distance* between prior and posterior distributions. For two probability distributions $p(\boldsymbol{\theta})$ and $q(\boldsymbol{\theta})$, the KL divergence is defined as:

$$D_{KL}(p \parallel q) = \int p(\boldsymbol{\theta}) \log \frac{p(\boldsymbol{\theta})}{q(\boldsymbol{\theta})} d\boldsymbol{\theta} \tag{17}$$

For the specific case of measuring information gain from observations, we calculate the KL divergence between the posterior $p(\boldsymbol{\theta}|\boldsymbol{y})$ and prior $p(\boldsymbol{\theta})$ distributions:

$$D_{KL}(p(\boldsymbol{\theta}|\boldsymbol{y}) \parallel p(\boldsymbol{\theta})) = \int p(\boldsymbol{\theta}|\boldsymbol{y}) \log \frac{p(\boldsymbol{\theta}|\boldsymbol{y})}{p(\boldsymbol{\theta})} d\boldsymbol{\theta} \tag{18}$$

A larger KL divergence indicates a greater difference between our posterior and prior beliefs. In information theory, this divergence quantifies the additional nats (or bits, when using logarithms with base 2) needed to represent the posterior relative to the prior. In contrast, a KL divergence close to zero suggests that the data added little information beyond what was already known from the prior.

Since the KL divergence is generally intractable for complex distributions, we can approximate it using Monte Carlo integration [33]. VBMC provides direct estimates of the posterior densities through its variational approximation, making this computation straightforward. Given samples $\{\boldsymbol{\theta}^{(i)}\}_{i=1}^{N}$ from the posterior distribution, we can estimate:

$$D_{KL}(p(\boldsymbol{\theta}|\boldsymbol{y}) \parallel p(\boldsymbol{\theta})) \approx \frac{1}{N} \sum_{i=1}^{N} \log \frac{p(\boldsymbol{\theta}^{(i)}|\boldsymbol{y})}{p(\boldsymbol{\theta}^{(i)})} \tag{19}$$

3. Case study overview: Bridge 705 in Amsterdam

3.1. Bridge description

The case study concerns bridge 705, a reinforced concrete slab bridge located in Amsterdam, as shown in Fig. 1. The bridge was built in 1960 and up to the measurement campaign in 2018 no visible

Fig. 1. Side view of bridge 705.
Source: Picture taken from [34].

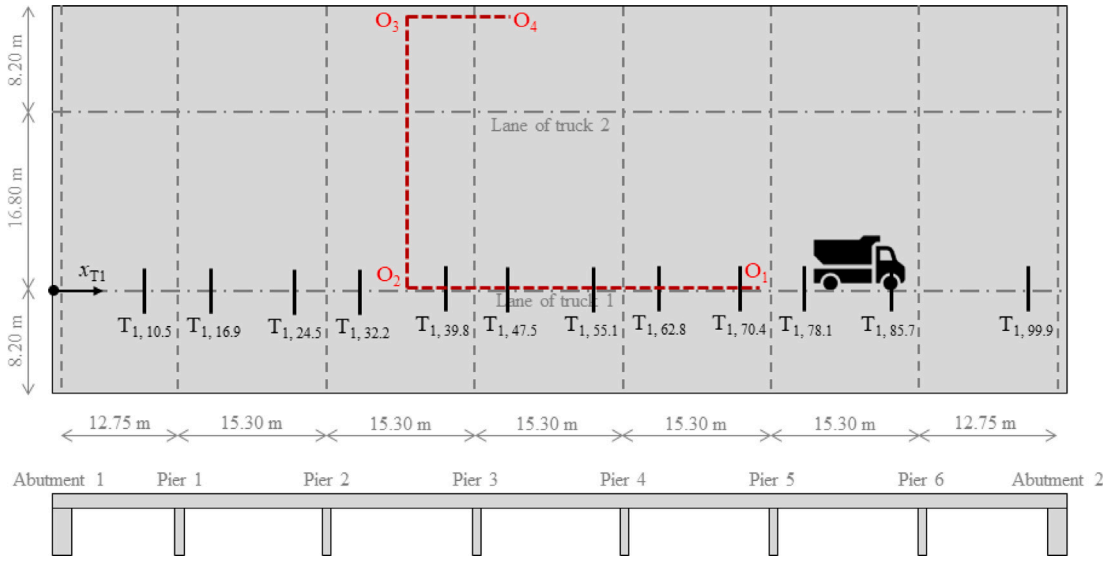


Fig. 2. Illustration of the moving truck 1 (T_1) at twelve positions on bridge 705 that was used during the measurement campaign and the location of the optic fibre strain sensor (red dashed line). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
Source: Picture adapted from [34].

damage or signs of corrosion were detected. The bridge deck spans 102 m in length and 33.2 m in width, and is supported at each end by abutments. In addition, six rows of four V-shaped piers provide intermediate support. Each of these 24 piers is monolithically connected to the deck at its top. The piers at the two central rows have hinged supports at their base, while the rest use rollers. The bridge accommodates two tram tracks, driving lanes and sidewalks.

3.2. Measurement campaign

At the end of 2018, bridge 705 was instrumented and diagnostically loaded. In total, about 90 sensors were used that recorded strains, translations, accelerations, and temperature. During two nights, the bridge was closed for traffic and controlled static and dynamic load tests were executed using two sand trucks. One of the main objectives of this extensive, on-site measurement campaign was to compare the information content of structural response measurements obtained from traditional and innovative sensor technologies. More details of this measurement campaign, the used sensors, the different load tests, and the data processing are presented in [34].

In this paper, we consider the strain data that was acquired during the static load test series T_1 with a single truck (truck 1) using a high resolution optic fibre sensor. Truck 1 has four axles and a total mass of 34.95 tonnes. Fig. 2 presents the twelve front axle positions from the left abutment (x_{T1}) in this test series, with x_{T1} values equal to 10.5, 16.9, 24.5, 32.2, 39.8, 47.5, 55.1, 62.8, 70.4, 78.1, 85.7 and 99.9 m. Fig. 2 further indicates the location of the optic fibre by the red dashed line. This distributed optic fibre sensor (DFOS) was placed in a milled notch and bonded by an epoxy adhesive over a total length of about 70 m along the bottom side of the bridge deck. It has a 10 cm distance between consecutive measuring points, where at each measuring point the strain value has been averaged over a spatial range of 20 cm. The reader is referred to Wosniok et al. [35] for further information about the DFOS configuration. In this paper, we only use the data that is measured over the about 35 m long distance between the points O_1 and O_2, since this part of the fibre is right below the lane

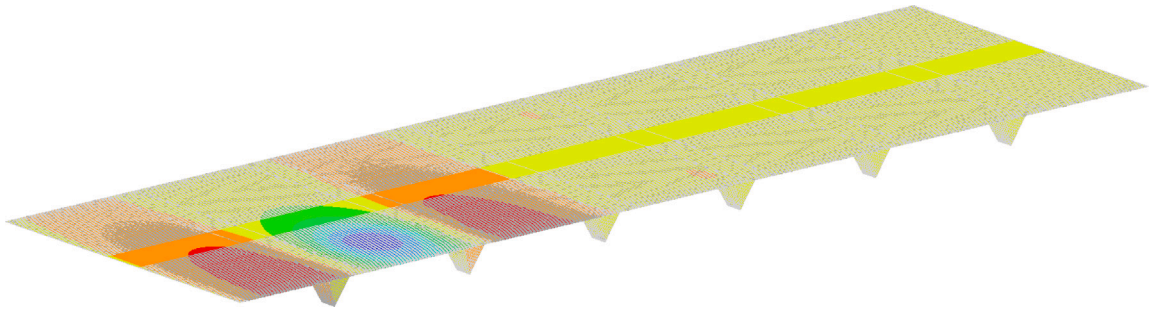


Fig. 3. Overview of the three-dimensional finite element model under a concentrated load in the second span.
Source: Picture taken from [34].

of truck 1. Some data processing was needed to compensate for small temperature changes that occurred during the test series, and to apply offset adjustments. For more details on these corrections, see [36].

3.3. Structural model

The strain responses of bridge 705 at the location of the DFOS are simulated using a three-dimensional finite element model (see Fig. 3). The deck and V-shaped piers are modelled with 68,649 quadrilateral eight-node shell elements, accounting for all thickness variations. With 207,848 nodes, each having three translational and two rotational degrees of freedom, the finite element model results in over 1 million degrees of freedom. Given the relatively low load intensities during the test series, nonlinearity in the concrete material behaviour can be neglected. Instead, the concrete is modelled as a linear–elastic material, for which the modulus of elasticity E_c is one of the two physical parameters that will be estimated. The concrete is assumed to be homogeneous and to have a constant value over the entire bridge deck. The finite element model accounts for contributions of the pavement and tramway rails to the deck stiffness by modification of the modulus of elasticity in the corresponding zones. The abutments on both sides of the deck are not explicitly modelled. Instead, vertical supports are placed at the bearing locations, and nodal supports are applied at the bottom of the piers, each with (i) horizontal and vertical constraints on the translational degrees of freedom, and (ii) a spring element for one of the rotational degree of freedom, in line with the bridge's structural design. The rotational spring stiffness K_r is the second physical parameter that will be estimated. The only modelled load is the weight of truck 1, where each of its wheel prints is modelled by a uniformly distributed load over a finite area. The movement of the truck is simulated with static analyses in discrete steps of one metre. All analyses are performed using the finite element software DIANA FEA. The accuracy of this model was validated in [34], where the finite element model with parameter values chosen a priori showed good agreement with discrete strain field measurements, achieving an R^2 score of 0.84. This demonstrates that the model reasonably captures the bridge behaviour under static loads, while leaving room for further refinement through parameter estimation.

The evaluation of the finite element model takes between 10 and 13 min. Since a Bayesian parameter estimation can require many of these evaluations, it may result in an excessively high wall clock time. As mentioned in Section 1, this computational challenge can be tackled by using Bayesian methods that need less likelihood function evaluations or by replacing the structural model responses with surrogate models that are multiple orders of magnitude faster to evaluate. In this paper, both strategies are adopted because we aim to solve multiple parameter estimation tasks.

The surrogate models replace the strain responses from the finite element model at the location of the DFOS. Between the points O_1 and O_2, the finite element model has 336 element nodes, so 336 surrogates are needed for each of the twelve truck positions that we consider. The surrogates are constructed with Gaussian processes (GPs). All GPs share the same hyperparameters, which are estimated by optimizing the marginal likelihood. Each GP consists of a two-dimensional input space that is formed by the two physical parameters to be estimated, E_c and K_r. Due to the low dimensionality of the input space, a points grid is used over which the finite element model was evaluated and then the GP models are fitted. The used grid has 112 points: 7 values between 10 to 100 GPa for E_c and 16 values between 10^0 to 10^{12} N-mm/rad for K_r. A finer discretization is used for K_r because it spans several orders of magnitude and has a more nonlinear influence on the structural response compared to E_c.

After fitting, the surrogate model predictions are compared against left out finite element model results. Since the surrogates deliver responses virtually identical to those from the finite element analyses, the 112 points provide sufficient information to construct an accurate surrogate for the finite element model. Considering this, there is no downside to using the surrogates in our parameter estimation tasks. Nevertheless, the proposed methodology with VBMC is equally applicable if a finite element model is used directly.

For additional information about the finite element model and its surrogates, the reader is referred to Rózsás et al. [34].

3.4. Probabilistic model

We adopt the data generating process from Eq. (7), but adding a multiplicative term ρ on the physical model ν:

$$y_i = \rho \cdot \eta(x_i, \boldsymbol{\theta}) + \delta(x_i) + \epsilon_i, \quad i = 1, \ldots, n \tag{20}$$

The term $\rho \sim \mathcal{N}(1, \sigma_m)$ is an i.i.d. variable. This signals that part of the model discrepancy might also scale with the magnitude of the model predictions. The modification was deemed necessary for the case study after observing that the model predictions were smooth, while measurements in the more strained areas exhibited fast-varying peaks. While the exact cause of these discrepancies—whether due to sensor errors or other factors—remains uncertain, for simplicity, we assume the model discrepancy contributes to these variations and therefore modify Eq. (10) accordingly:

$$\boldsymbol{y} \sim \mathcal{N}(\eta(\boldsymbol{x}, \boldsymbol{\theta}), \sigma_m^2 \mathrm{diag}(\eta(\boldsymbol{x}, \boldsymbol{\theta}))^2 + \boldsymbol{K} + \sigma_\epsilon^2 \boldsymbol{I}) \tag{21}$$

where $\mathrm{diag}(\eta(\boldsymbol{x}, \boldsymbol{\theta}))^2$ is a diagonal matrix composed of the squared values of $\eta(\boldsymbol{x}, \boldsymbol{\theta})$. We do not infer σ_m, but instead, we treat it as a fixed parameter. We assume $\sigma_m = 0.05$, which provides a balance between trusting our physical model while still allowing for some room for multiplicative errors. Additionally, 0.05 is a typical value in structural probabilistic modelling literature, being recommended in the JCSS Probabilistic Model Code - Part 3.9: Model uncertainties, for stresses in finite element models [37].

We also fix the standard deviation of the measurement error. This is possible because we have DFOS measurements taken before the application of the truck loads. In this situation, the true variations in

Table 2
Prior distribution of physical and statistical parameters.

Parameter	Units	Distribution	Hyperparameters
E_c	GPa	Uniform	$a = 10, b = 100$
$\log_{10} K_r$	N-mm/rad	Uniform	$a = 0, b = 12$
σ_a	microstrains	Gamma	$\alpha = 2, \beta = 2$
l	m	Gamma	$\alpha = 3, \beta = 1$

Table 3
Description of sub-cases used for evaluating the performance of VBMC. For the load cases, see Fig. 2.

Sub-case	Covariance function	Load cases	Method
A1	Matérn, $\nu = 1.5$	$T_{1,39.8}$	MCMC
A2			VBMC

Table 4
Hard bounds and plausible bounds for the VBMC algorithm.

Parameter	Units	Hard bounds		Plausible bounds	
		Lower	Upper	Lower	Upper
E_c	GPa	10.0	100.0	20.0	60.0
$\log_{10} K_r$	N-mm/rad	0.0	12.0	1.0	11.0
σ_a	microstrains	0.0	5.0	0.2	2.0
l	m	0.0	10.0	0.2	5.0

strains are zero, so we can attribute the fluctuations to noise, resulting in a fitted value of σ_ϵ of 0.1 microstrains after rounding.

In contrast, we infer the statistical parameters of the covariance function σ_a and l, along with the physical parameters E_c and $\log_{10} K_r$. Table 2 shows the chosen prior distributions. For the physical parameters, we have opted for wide uniform distributions, allowing the data to drive the inference process. On the other hand, we use Gamma distributions for the statistical parameters, which are appropriate for strictly positive variables. However, we slightly restrict these distributions, recognizing that they require some control to ensure meaningful results as the inference is sensitive to them.

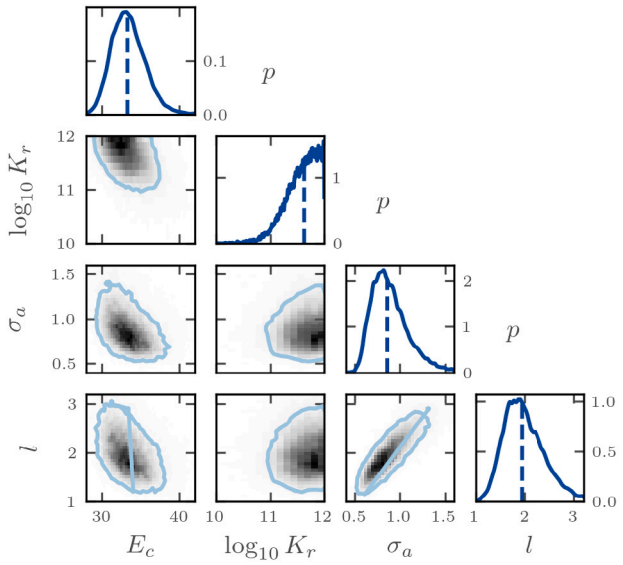
While the parameters σ_m and σ_ϵ could also be inferred, doing so might introduce excessive flexibility in the statistical model that could confound with the physical model, potentially leading to less meaningful results. For a broader discussion about this topic, see [25].

4. Performance of variational Bayesian Monte Carlo

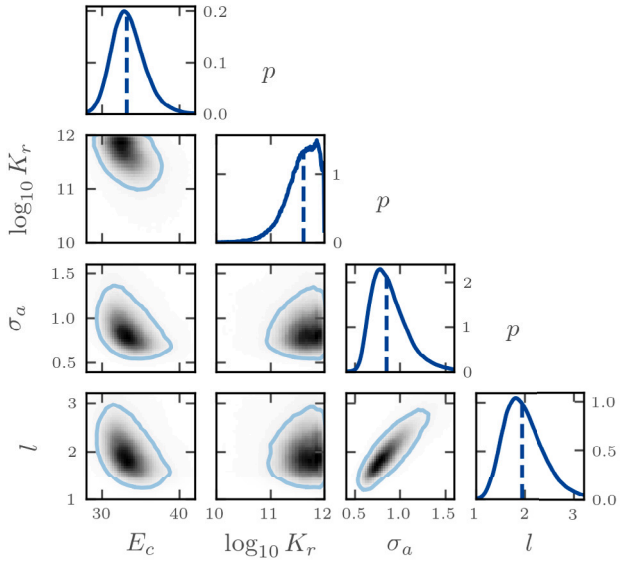
The use of VBMC allows to obtain accurate posterior distributions with a limited number of likelihood evaluations. From Section 3.3, we have cheap-to-compute Gaussian process surrogate models available for the finite element model of the bridge at DFOS locations. This enables the benchmarking of VBMC against a more established Bayesian inference method. We use the Affine invariant ensemble sampler MCMC method [38] as implemented in the `emcee` Python package [39]. A description of the sub-cases used for the comparison is shown in Table 3.

We run sub-case A1 using the default settings of `PyVBMC`. The algorithm requires the user to define both hard and plausible bounds. Table 4 provides the selected bounds. For sub-case A2, we use the default settings of `emcee`. In this case, the user must specify the number of walkers and steps. We choose 40 walkers, with 2000 steps. The initial positions of the walkers are sampled from the prior distribution. Notably, only one load case is used for inference, as incorporating additional load cases would significantly slow down the MCMC run. This is due to the computational complexity of Gaussian likelihood calculation, which scales cubically with the number of datapoints.

The summary statistics of the posterior distributions for both cases are presented in Table 5. Across all parameters, the mean and standard deviations produced by VBMC closely align with those obtained through MCMC. Additionally, the 1D and 2D marginal distributions



(a) Sub-case A1 – MCMC



(b) Sub-case A2 – VBMC

Fig. 4. Corner plot of the posterior distribution. The diagonal entries correspond to the 1D marginal distributions for E_c (GPa), $\log_{10} K_r$ (N-mm/rad), σ_a (microstrains), and l (m). The rest of the entries are 2D marginal distributions.

exhibit similar shapes, as illustrated in Fig. 4. It is worth noting that VBMC shows smoother densities due to the algorithm use of a parametric approximation of the posterior.

We have verified that VBMC provides an accurate posterior estimation with respect to MCMC. To evaluate its efficiency, we compare the convergence of parameter estimates for each method. As shown in Figs. 5 and 6, VBMC achieves a significant reduction in computational cost, converging after only 145 likelihood function evaluations compared to approximately 30,000 evaluations needed by MCMC. This represents a reduction in the number of evaluations by a factor of around 200. As described in Section 3.3, each FE model evaluation takes around 10 min, which would make running MCMC directly on the FE model

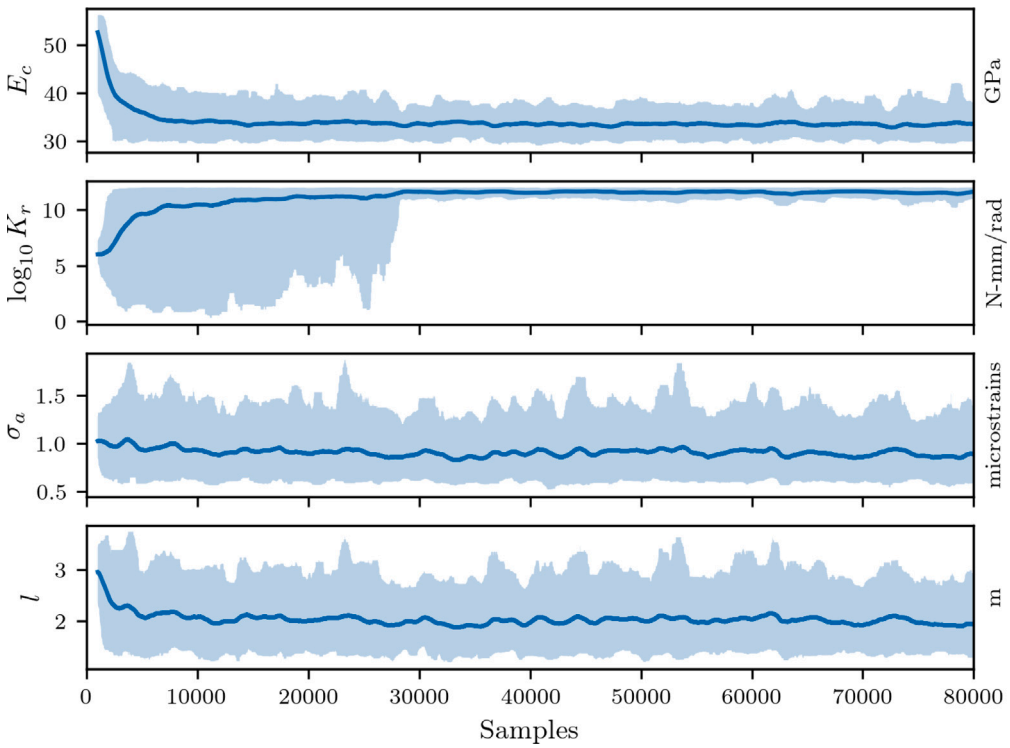


Fig. 5. Evolution of MCMC parameter estimates. The solid line and the shadowed area correspond to the mean and 95% credible interval, calculated using only the previous 1000 samples. Results stabilize around 30,000 samples.

Table 5
Comparison of 1D marginal posterior parameter estimates.

Parameter	Unit	Sub-case A1		Sub-case A2	
		Mean	std. dev.	Mean	std. dev.
E_c	GPa	33.42	2.20	33.42	2.11
$\log_{10} K_r$	N-mm/rad	11.56	0.42	11.57	0.30
σ_a	microstrains	0.90	0.22	0.89	0.20
l	m	2.02	0.44	1.99	0.40

Table 6
Considered covariance functions.

Function	$k(\mathbf{x}, \mathbf{x}')$
Independent	$\sigma^2 I(\mathbf{x} = \mathbf{x}')$
Exponential	$\sigma^2 \exp\left(-\frac{\lVert \mathbf{x}-\mathbf{x}' \rVert}{l}\right)$
Matérn, $\nu = 1.5$	$\sigma^2 \left(1 + \frac{\sqrt{3}\lVert \mathbf{x}-\mathbf{x}' \rVert}{l}\right) \exp\left(-\frac{\sqrt{3}\lVert \mathbf{x}-\mathbf{x}' \rVert}{l}\right)$
Matérn, $\nu = 2.5$	$\sigma^2 \left(1 + \frac{\sqrt{5}\lVert \mathbf{x}-\mathbf{x}' \rVert}{l} + \frac{5\lVert \mathbf{x}-\mathbf{x}' \rVert^2}{3l^2}\right) \exp\left(-\frac{\sqrt{5}\lVert \mathbf{x}-\mathbf{x}' \rVert}{l}\right)$
Squared Exponential	$\sigma^2 \exp\left(-\left(\frac{\lVert \mathbf{x}-\mathbf{x}' \rVert}{l}\right)^2\right)$

infeasible, as it would require around 5000 computing hours, whereas running VBMC directly on the FE model would require only about 24 computing hours. Hamiltonian Monte Carlo methods such as the No-U-Turn Sampler [40] can be more efficient, but require gradients of the posterior density, which are not accessible with a black-box physical model, and are still likely to need a few thousand evaluations.

In the remainder of this section, we demonstrate the effectiveness of the procedure outlined in Section 2.4.1 for adapting VBMC results to obtain posterior predictive estimates. Since we have stored the 145 parameter samples queried by VBMC and the simulations responses of interest, we can utilize these input–output pairs to build surrogate models. Specifically, we employ a zero-mean Gaussian process for each output dimension, utilizing shared kernel hyperparameters that are optimized by maximizing the marginal likelihood. This is possible because all output dimensions are of the same type: they represent strains of similar magnitude and have a similar sensitivity to the input parameters.

Since VBMC prioritizes querying points that highly contribute to the accuracy of the posterior distribution, it follows that a surrogate model of the structural responses built on these points will also exhibit high accuracy within the posterior region. To illustrate this, we compare the performance of the surrogate model mean predictions against the real predictions at samples located on the boundary of the 0.99 credible region and at samples outside this region. As shown in Fig. 7, the surrogate model maintains excellent accuracy even on the boundary of the 0.99 credible region; however, its performance deteriorates outside this region. Nonetheless, because the contribution of samples outside the 0.99 credible region to the posterior predictive is minimal, the surrogate model remains sufficiently accurate for generating posterior predictive distributions.

5. Evaluation of covariance models

We have proposed a data generating process that accounts for the spatial dependency of model errors. For this, we need to specify the covariance function. As discussed in Section 2.3, this function should be distance-based and stationary; however, several functions meet these criteria. In this section, we evaluate various covariance models and explore how to select the most suitable one. This involves the use of posterior predictions and Bayesian model selection presented in Sections 2.4.1 and 2.4.2, respectively.

Table 6 shows the covariance functions considered in this paper, where I is the indicator function, σ is the scale term and l is the correlation length. We selected these functions to represent a spectrum of decay properties, allowing us to investigate how different spatial correlation assumptions affect the model uncertainty. As we move down the table, the realizations of model discrepancy become progressively smoother.

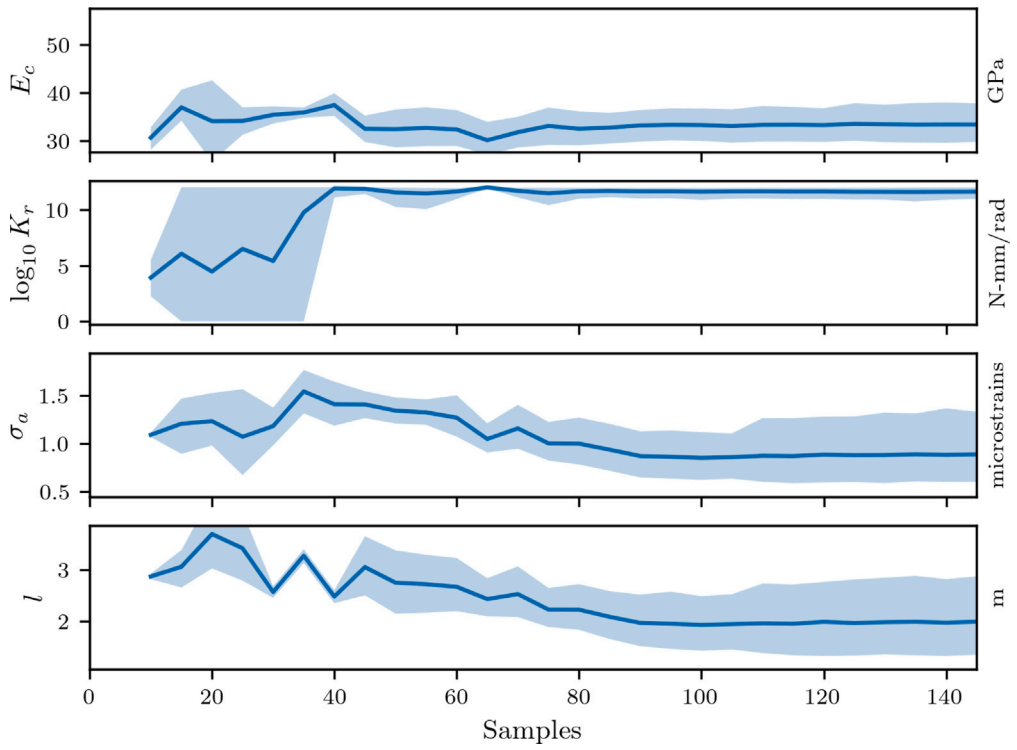
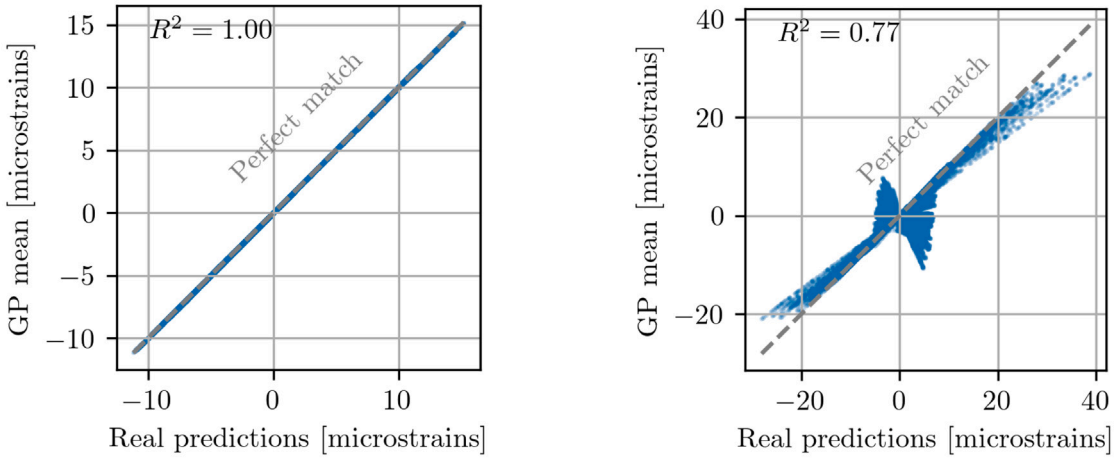


Fig. 6. Evolution of VBMC parameter estimates. The solid line and the shadowed area correspond to the mean and 95% credible interval, calculated using the current variational approximation for that step. Convergence is obtained at 145 samples.



(a) Samples on 0.99 credible region boundary

(b) Samples outside the 0.99 credible region

Fig. 7. Comparison of the real predictions vs. the mean predictions of the surrogates trained with VBMC queried samples. The accuracy is excellent even on the boundary of the 0.99 credible region. The surrogate performance deteriorates outside the 0.99 credible region, but the contribution of these samples to the posterior predictive is minimal.

Table 7
Description of sub-cases used for evaluating the covariance functions. For the load cases, see Fig. 2.

Sub-case	Covariance function	Load cases	Method
B1	Independent		
B2	Exponential		
B3	Matérn, $\nu = 1.5$	$T_{1,39.8}$	VBMC
B4	Matérn, $\nu = 2.5$		
B5	Squared exponential		

To investigate which of these functions is the most appropriate, we use the sub-cases shown in Table 7. For the inference process, we use data from only one load case and reserve the remaining load cases for posterior predictive validation. This allows us to evaluate how well the inferred models generalize to unseen data. Inference was conducted using the same VBMC settings detailed in Section 4. The number of likelihood evaluations required for each sub-case ranged between 100 and 200. A summary of the resulting posterior distributions is provided in Table 8.

In general, the obtained posterior credible intervals are considerably narrower than the prior credible intervals. We also observe

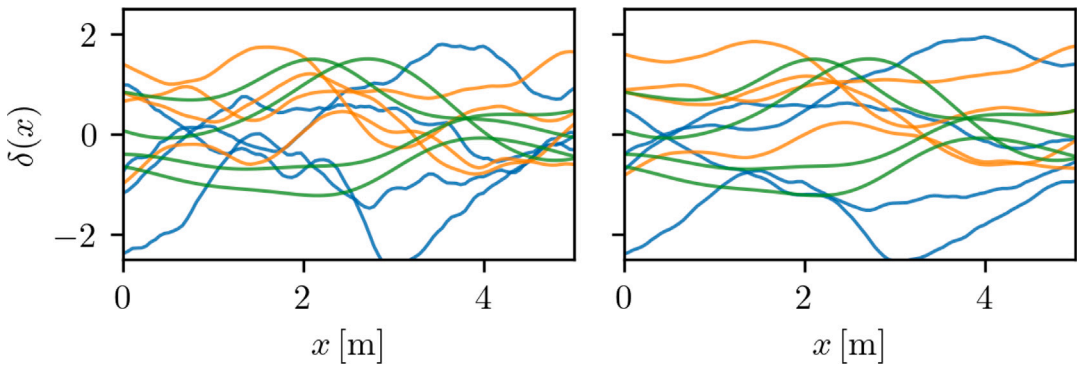


Fig. 8. Realizations of model discrepancy functions $\delta(x)$ for the covariance functions. The blue, orange and green lines correspond to sub-cases B3, B4 and B5, respectively. Left: Using the same hyperparameters for all three functions. Right: Using inferred mean posterior values of σ_a and l. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 8
Summary of 1D marginal posterior parameter estimates. The value at the top of each cell corresponds to the mean. The values inside the brackets are the 2.5% and 97.5% percentiles.

Sub-case	E_c [GPa]	$\log_{10} K_r$ [N-mm/rad]	σ_a [microstrains]	l [m]
B1	45.28 [42.80, 47.76]	11.19 [10.87, 11.49]	0.64 [0.59, 0.69]	–
B2	40.51 [35.46, 46.38]	10.63 [2.58, 11.92]	0.62 [0.47, 0.81]	4.41 [2.44, 7.11]
B3	33.42 [29.74, 38.11]	11.59 [10.89, 11.98]	0.88 [0.59, 1.37]	1.99 [1.33, 2.92]
B4	32.51 [29.08, 36.73]	11.59 [10.92, 11.98]	0.88 [0.61, 1.30]	1.53 [1.16, 1.99]
B5	32.60 [29.24, 36.87]	11.58 [10.87, 11.99]	0.79 [0.55, 1.11]	0.98 [0.78, 1.17]

Table 9
Strength of evidence for the studied sub-cases. Each entry ij represents how strongly the model in row i is favoured over the model in column j. Models B1 and B3 are the least and most favoured, respectively.

Sub-case	B1	B2	B3	B4	B5
B1	–	Negative	Negative	Negative	Negative
B2	Decisive	–	Negative	Negative	Negative
B3	Decisive	Decisive	–	Barely mention	Substantial
B4	Decisive	Decisive	Negative	–	Substantial
B5	Decisive	Decisive	Negative	Negative	–

Table 10
Percentage of measurements within the 95% credible interval of the posterior predictive distribution. While all sub-cases show an adequate fit to the data, B3 and B4 favoured for unobserved data.

Sub-case	Inference data	Unobserved data
B1	96.19%	75.81%
B2	96.77%	78.30%
B3	96.48%	89.81%
B4	96.48%	88.64%
B5	96.48%	85.12%

relatively large differences in the inferred values of E_c across the different models. Specifically, models B1 and B2 exhibit higher values for E_c compared to models B3, B4, and B5, which show more similar and lower values. Additionally, all cases show inferred $\log_{10} K_r$ parameters that are closer to their prior upper bound. This indicates that the bottom supports of the pillars behave as clamped rather than as hinged, a finding that aligns with the observations made by Rózsás et al. [34] in their study of this specific bridge.

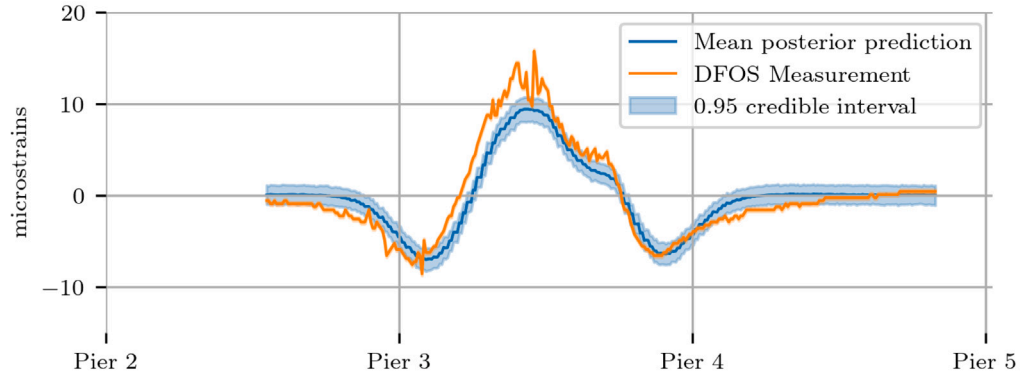
Moreover, the inference results reveal a modest amount of model discrepancy, with mean σ_a values remaining below 1.0 microstrains across all cases. This is an encouraging outcome, suggesting that the finite element model used is of high quality. We also observe a trend where the correlation length l decreases as the smoothness of the covariance function increases. To understand the implications of this behaviour, we inspect random realizations of the model discrepancy function $\delta(x)$ for models B3, B4, and B5, using the posterior mean values of σ_a and l (see Fig. 8). Despite the covariance functions having different decay properties, these realizations exhibit similar trends. This suggests that the decrease in correlation length compensates for the increase in smoothness, allowing the models to fit the data similarly. Therefore, the specific choice of covariance function may not be critical, provided that the correlation parameters are also inferred.

Although VBMC does not directly compute the evidence, it provides a proxy for it: the evidence lower bound (ELBO). We use it to calculate Bayes factors for each pair of models and interpret the strength of evidence according to Jeffreys [32]. The results are presented in Table 9. It is reassuring that sub-case B1 is the least favoured by Bayesian model selection, despite being the most parsimonious, i.e., it has fewer inferred parameters than the rest. This aspect is usually weighted favourably in Bayesian model selection. The assumption behind sub-case B1 was that model discrepancy occurred uncorrelated to the spatial proximity of points. It is easy to determine why this intuition was flawed by considering that both the true physical strain responses and the simulation are expected to be smooth, so their difference should be smooth as well. Sub-cases B3 and B4, corresponding to moderately smooth Matérn covariance functions, yield the best results.

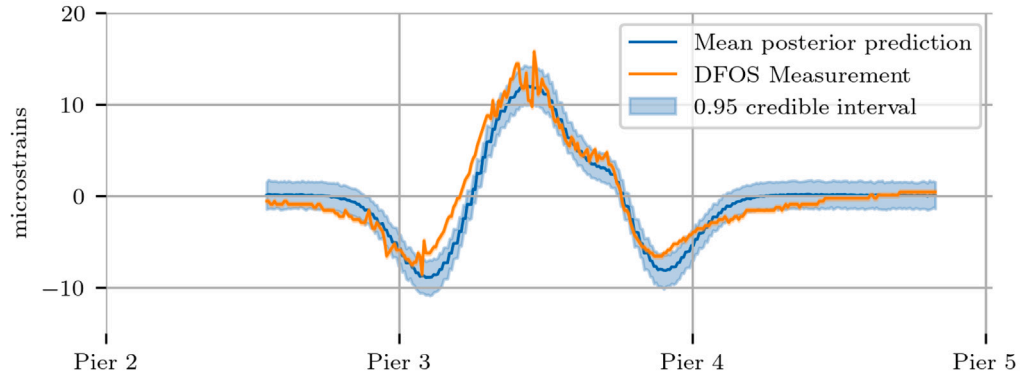
Models can also be compared by their predictive capacity using the inferred posterior parameters. We compute the posterior predictive distribution, as outlined in Section 2.4.1, for both the observed load case $T_{1,39.8}$, and the unobserved $T_{1,47.5}$, $T_{1,55.1}$, $T_{1,62.8}$, and $T_{1,70.4}$. As a metric, we use the percentage of measurements that fall within the 95% credible interval of the posterior predictive distribution. The results are shown in Table 10.

We observe a nearly uniform percentage across all models for the inference data, indicating that they all perform similarly well in fitting the observed data. However, the results for the load cases not used for inference present a different scenario, with models B3 and B4, which were most favoured by Bayesian model selection, showing better performance. This suggests that Bayesian model selection may be effective in identifying models that generalize better to unobserved data, although it is worth noting that some researchers have reported contradictory findings in this area [41]. Additionally, these results demonstrate that a single load case was sufficient for reliable system identification, validating our initial choice.

Since we are interested in spatial correlations, it is useful to show the posterior predictions across the length of the DFOS. In Figs. 9 and 10, we show fixed-load plots for load cases $T_{1,55.1}$ and $T_{1,70.4}$, which were not used for inference. The measurements in orange show the fast-varying peaks in the high strained areas, as described in Section 3.4.



(a) Sub-case B1



(b) Sub-case B3

Fig. 9. Comparison of longitudinal static strain measurements along the bottom face of the bridge deck for load case $T_{1,55.1}$, which was not used for inference. The plots show DFOS measurements (orange) versus mean posterior predictions (blue) with 95% credible intervals (shaded blue). Sub-case B3 shows better agreement between predictions and measured strains than sub-case B1. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 11
Description of sub-cases used for evaluating inference across different resolutions. For the load cases, see Fig. 2.

Sub-case	Res.	Equivalence	Cov. function	Load cases	Method
C1	0.1 m	DFOS	Matérn, $\nu = 1.5$	$T_{1,all}$, except $T_{1,70.4}$	VBMC
C2	0.2 m	DFOS			
C3	0.5 m	FBG			
C4	1.0 m	FBG			
C5	2.0 m	Strain gauges			
C6	4.0 m	Strain gauges			

There is moderate agreement between the mean predictions, represented by the blue lines, and the measurements for sub-case B1, while a very good agreement is observed for sub-case B3. Moreover, the credible interval, shown as a shadowed area, encompasses most of the measurements for sub-case B3, which is not the case for sub-case B1. Both observations are consistent with the results presented in Table 10.

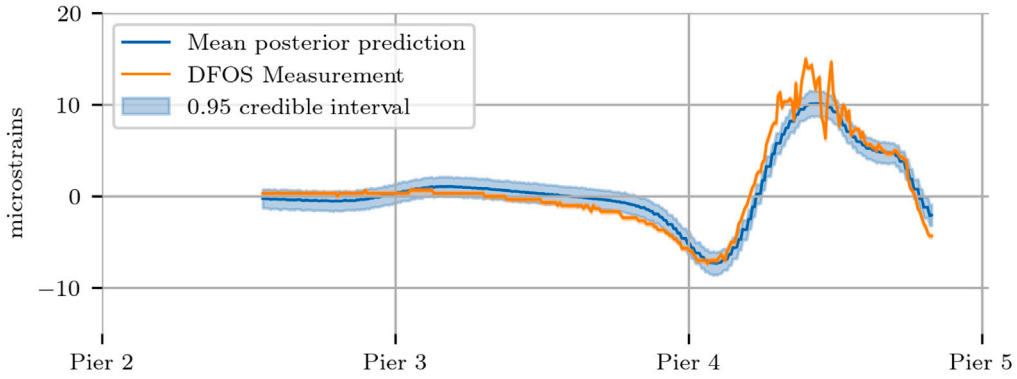
6. Value of distributed fibre optic sensor

In the previous sections, we developed an efficient approach to perform Bayesian system identification using fibre optic data, extended VBMC for posterior predictive estimation, and examined various covariance functions to account for spatial correlations. However, a critical question is whether high-resolution optic fibre data, despite potential high correlations, provides greater value in a Bayesian context. To evaluate this, we consider six sub-cases with different resolutions, as detailed in Table 11.

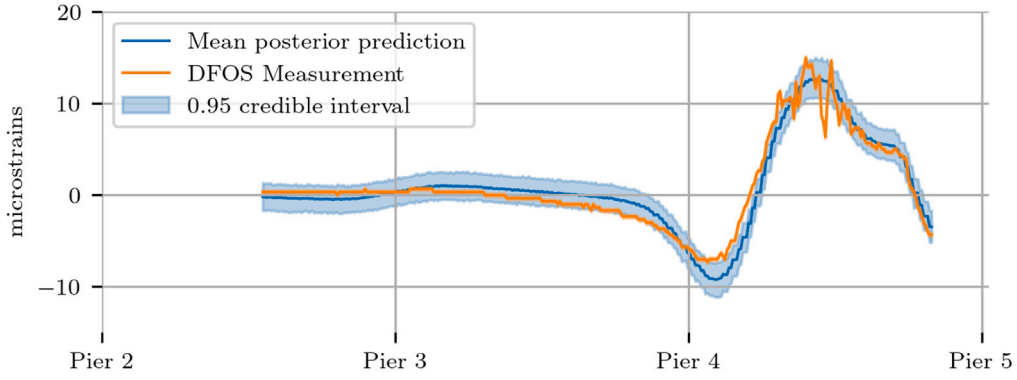
The base sub-case C1 uses the full resolution of the data, with strains every 0.1 m along the fibre. For the other sub-cases, we downsample the data to match the respective resolutions by selecting data points at regular intervals corresponding to the desired spacing, starting from a fixed reference point. For example, in C3 with a resolution of 0.5 m, we select every fifth data point from C1, resulting in measurements at positions 0 m, 0.5 m, 1.0 m, and so on. We interpret these resolutions as representative of different strain sensing technologies, such as distributed optic fibre sensors (DFOS), Fibre-Bragg gratings (FBG) and traditional strain gauges. To ensure that each sub-case has sufficient data for robust Bayesian inference, particularly in the lower-resolution cases, we include measurements from all available load cases ($T_{1,all}$), with the exception of $T_{1,70.4}$, which we reserve for posterior predictive checks.

It is well-known that the posterior distribution converges at a rate proportional to $n^{-1/2}$, where n is the number of independent and identically distributed (i.i.d.) observations [42]. However, in our case, the assumption of independence is violated due to the presence of a correlated model discrepancy term. This raises uncertainty about whether the posterior will contract as the resolution increases. To investigate this, we calculate the 95% credible intervals for the marginal posterior parameters, as shown in Fig. 11. We observe that the posterior contracts with more data points, even after accounting for spatial correlations. This indicates that the additional observations from the higher resolution are not redundant but provide extra value in a Bayesian context.

Apart from comparing 1D-marginal credible intervals, the information content for different resolutions can also be assessed using the



(a) Sub-case B1



(b) Sub-case B3

Fig. 10. Comparison of longitudinal static strain measurements along the bottom face of the bridge deck for load case $T_{1,70.4}$, which was not used for inference. The plots show DFOS measurements (orange) versus mean posterior predictions (blue) with 95% credible intervals (shaded blue). Sub-case B3 shows better agreement between predictions and measured strains than sub-case B1. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 12
KL divergence between posterior and prior distribution for different fibre resolutions. The values were calculated with a Monte Carlo approach using 10,000 samples.

Sub-case	D_{KL}
C1	12.66
C2	11.45
C3	10.11
C4	9.68
C5	8.83
C6	7.28

Table 13
Performance of posterior predictions for load case $T_{1,70.4}$, which was not used for inference. The coefficient of determination (R^2) and mean squared error (MSE) were calculated between the measurements and mean posterior predictions. The mean width corresponds to the 95% equal-tailed posterior predictive credible interval.

Sub-case	R^2	MSE	Mean C.I. width
C1	0.93	1.58	2.38
C2	0.93	1.70	2.56
C3	0.93	1.71	2.52
C4	0.92	1.63	2.56
C5	0.92	1.57	2.55
C6	0.92	1.63	2.44

Kullback–Leibler (KL) divergence. Table 12 shows the KL divergence for different fibre resolutions. These values were calculated with the Monte Carlo approach described in Section 2.4.3, using 10,000 samples. The results indicate that the KL divergence increases with higher resolutions, reflecting a greater information gain as resolution improves, even when correlations are present.

Despite the benefits of higher resolution in terms of posterior concentration and information content, the impact on predictive performance is more limited. As shown in Table 13, all sub-cases show consistent agreement between measurements and mean posterior predictions, as quantified by the coefficient of determination and mean squared error. Additionally, the average width of the 95% posterior predictive credible intervals remains stable across all cases. This suggests that while finer discretization may provide more detailed posterior distributions, it does not necessarily translate into substantially improved predictive capabilities for this particular load case.

7. Discussion

The use of VBMC in our case study resulted in a dramatic reduction in computational costs compared to MCMC while maintaining accuracy. This efficiency comes from the ability of VBMC to build a surrogate of the log joint probability using relatively few evaluations of the physical model, after which inference can proceed using the cheap surrogate. Performance remained stable across all sub-cases, requiring minimal fine-tuning. This characteristic is particularly valuable for practitioners who may not be experts in Bayesian inference. Moreover, the method does not impose hard restrictions, offering flexibility for inference across diverse scenarios. For these reasons, the application of VBMC for general use in structural engineering shows great potential.

Despite these advantages, VBMC has limitations that deserve consideration. Huggins et al. [28] noted that VBMC might struggle with

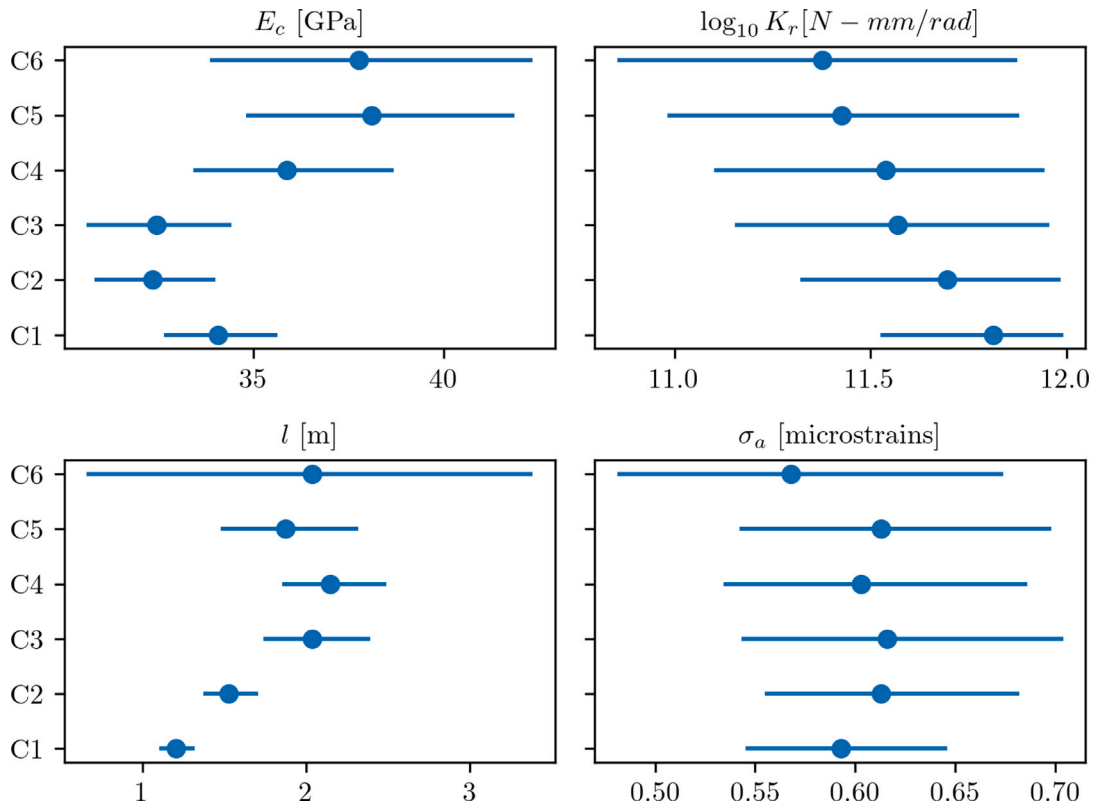


Fig. 11. Concentration of parameter posteriors for different fibre resolutions. With increasing resolution, the credible intervals become narrower.

cases involving more than 20 parameters. Although engineering models tend to have moderate dimensionality, some applications can exceed this threshold significantly, such as when discretizations of random fields are involved. Additionally, our proposed method for obtaining posterior predictive distributions may face challenges with highly non-linear physical models where the surrogate approximation could be less accurate. Furthermore, Igea and Cicirello [43] found that VBMC failed in the presence of highly multi-modal posteriors and proposed an extension to address this problem.

Regarding correlations, our choice to represent model discrepancy with a Gaussian process was partially motivated by convenience, as it enabled analytical calculation of the likelihood. Nevertheless, this can be justified theoretically, as model discrepancy arises from numerous small effects not captured by the idealized physical model. By the Central Limit Theorem, the aggregate effect of these many independent sources tends towards a Gaussian distribution. Extending this reasoning spatially, the Gaussian process emerges as a natural choice for modelling the spatial distribution of these accumulated effects, as it represents the infinite-dimensional generalization of multivariate normal distributions. The importance of considering spatial correlations is clearly demonstrated in our results, where the model version that ignored correlations (sub-case B1 in Section 5) showed significantly worse performance in terms of both parameter inference and predictions, with Bayes factors strongly favouring all other correlation models.

On the other hand, selecting a specific covariance function remains challenging. In our case, assuming a stationary function with relatively smooth behaviour was justified, but other applications might require different approaches. An additional limitation is that, unlike physical parameters, prior distributions for covariance function parameters do not arise naturally, yet they can significantly impact inference. We recommend incorporating knowledge of the physical system whenever possible, potentially through physics-informed covariance functions.

The analysis of different sensor resolutions revealed clear advantages for using distributed fibre optic measurements when the primary objective is parameter inference, as evidenced by both narrower credible intervals and increased Kullback–Leibler divergence. This can be the case for monitoring systems where the parameters to be inferred are damage levels. Regarding predictive performance, the differences across resolutions were modest, but this should be interpreted in context: even the lowest-resolution case included measurements from multiple load cases, providing substantial information for inference. In applications where fewer load cases are available, the benefits of high-resolution measurements might become more pronounced for predictive tasks.

Our case study relies on fibre optic data, but the methodology is applicable to other forms of spatially dense measurements, such as those obtained from Digital Image Correlation. In this context, model discrepancy manifests in two or three dimensions, requiring the use of higher dimensional Gaussian processes. The choice between considering decoupled covariance functions for each dimension or a single function based on the three-dimensional distance between points will depend on the structure being modelled. For example, an orthotropic deck might need separate functions for each dimension, while a uniform slab could be better represented by a single function.

The widespread adoption of model-based approaches in Structural Health Monitoring (SHM) has been constrained, in part, by their significant computational demands. However, the efficiency of the proposed method could contribute to transition from purely data-driven to model-based strategies. This shift has the potential to offer more robust and interpretable results, as it combines the advantages of physics-based modelling with statistical inference.

8. Conclusions

This paper presents, to the authors' knowledge, the first application of Variational Bayesian Monte Carlo (VBMC) for Structural Health

Monitoring (SHM) in civil engineering structures. The study uses distributed fibre optic strain measurements from static load tests on a large reinforced concrete slab bridge to infer both structural and statistical model parameters. The following key conclusions are drawn:

1. VBMC requires only 100–200 likelihood function evaluations to successfully infer four parameters, significantly improving efficiency compared to traditional Markov Chain Monte Carlo methods, which typically require 10^3–10^5 evaluations.
2. Ignoring spatial correlation reduces model performance on test (hold-out) data: models with spatial correlation using the Matérn covariance function achieve 89% coverage for a 95% credible interval, while models assuming independence only reach 76%.
3. Increasing the spatial density of strain measurements along the optic fibre improves parameter estimates (narrower credible intervals, higher Kullback–Leibler divergence), but does not substantially enhance posterior predictions.

This work demonstrates that VBMC-based system identification, enriched by our proposed approach for calculating posterior predictive distributions, provides an efficient and accurate framework for SHM in real-world engineering structures. It also highlights the ability of the method to handle spatially dense measurement data, enabling the use of modern sensing technologies in SHM applications. Although based on a specific case study, these findings provide valuable insights for other engineering structures, advancing the practical application of Bayesian system identification in SHM.

CRediT authorship contribution statement

Andrés Martínez: Writing – original draft, Writing – review & editing, Visualization, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Arthur Slobbe:** Writing – review & editing, Writing – original draft, Supervision, Data curation, Conceptualization. **Árpád Rózsás:** Writing – review & editing, Writing – original draft, Supervision. **Iuri Rocha:** Writing – review & editing, Supervision. **Frans van der Meer:** Writing – review & editing, Supervision.

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

During the preparation of this work the authors used Claude Sonnet 3.5 in order to improve language and readability. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

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.

Acknowledgements

The authors would like to thank TNO Optics and BAM-Division 8.6 for their role in the collection of the distributed fibre optic measurement data used in the case study. Special thanks are also extended to the municipality of Amsterdam for facilitating the closure of the bridge over two nights. This research was partially conducted during the TNO Early Research Program-Structural Integrity, use case Concrete Bridge.

Data availability

Data will be made available on request.

References

[1] Bado MF, Casas JR. A review of recent distributed optical fiber sensors applications for civil engineering structural health monitoring. Sensors 2021;21(5):1818.
[2] Mousa MA, Yussof MM, Udi UJ, Nazri FM, Kamarudin MK, Parke GA, Assi LN, Ghahari SA. Application of digital image correlation in structural health monitoring of bridge infrastructures: A review. Infrastructures 2021;6(12):176.
[3] Daró P, La Mazza D, Longo M, Basone F, Chiariotti P, Cigada A, Mancini G. Continuous MEMS SHM systems to support RC bridges retrofitting interventions through an ongoing diagnosis. In: International symposium of the international federation for structural concrete. Springer; 2023, p. 1816–26.
[4] Dashti M, Stuart AM. The Bayesian approach to inverse problems. In: Handbook of uncertainty quantification. Cham: Springer International Publishing; 2017, p. 311–428, [Chapter 10].
[5] Ren W-X, Chen H-B. Finite element model updating in structural dynamics by using the response surface method. Eng Struct 2010;32(8):2455–65.
[6] Monchetti S, Viscardi C, Betti M, Clementi F. Comparison between Bayesian updating and approximate Bayesian computation for model identification of masonry towers through dynamic data. Bull Earthq Eng 2024;22(7):3491–509.
[7] Ierimonti L, Cavalagli N, Venanzi I, García-Macías E, Ubertini F. A transfer Bayesian learning methodology for structural health monitoring of monumental structures. Eng Struct 2021;247:113089.
[8] Standoli G, Salachoris GP, Masciotta MG, Clementi F. Modal-based FE model updating via genetic algorithms: Exploiting artificial intelligence to build realistic numerical models of historical structures. Constr Build Mater 2021;303:124393.
[9] Chai X, Rozsas A, Slobbe A, Teixeira A. Probabilistic parameter estimation and reliability assessment of a simulated sheet pile wall system. Comput Geotech 2022;142:104567.
[10] Salachoris GP, Standoli G, Betti M, Milani G, Clementi F. Evolutionary numerical model for cultural heritage structures via genetic algorithms: a case study in central Italy. Bull Earthq Eng 2024;22(7):3591–625.
[11] Pepi C, Gioffre' M, Grigoriu MD. Parameters identification of cable stayed footbridges using Bayesian inference. Meccanica 2019;54:1403–19.
[12] Kurent B, Friedman N, Ao WK, Brank B. Bayesian updating of tall timber building model using modal data. Eng Struct 2022;266:114570.
[13] Ramancha MK, Vega MA, Conte JP, Todd MD, Hu Z. Bayesian model updating with finite element vs surrogate models: Application to a miter gate structural system. Eng Struct 2022;272:114901.
[14] Sbarufatti C, Cadini F, Locatelli A, Giglio M, et al. Surrogate modelling for observation likelihood calculation in a particle filter framework for automated diagnosis and prognosis. In: Proceedings of the EWSHM 2018 9th European workshop on structural health monitoring (EWSHM 2018), Manchester, UK. 2018, p. 10–3.
[15] Torzoni M, Manzoni A, Mariani S. A deep neural network, multi-fidelity surrogate model approach for Bayesian model updating in SHM. In: European workshop on structural health monitoring. Springer; 2022, p. 1076–86.
[16] Li Q, Du X, Ni P, Han Q, Xu K, Yuan Z. Efficient Bayesian inference for finite element model updating with surrogate modeling techniques. J Civ Struct Heal Monit 2024;1–19.
[17] Zhou Y, Lu Z, Hu J, Hu Y. Surrogate modeling of high-dimensional problems via data-driven polynomial chaos expansions and sparse partial least square. Comput Methods Appl Mech Engrg 2020;364:112906.
[18] Titscher T, van Dijk T, Kadoke D, Robens-Radermacher A, Herrmann R, Unger JF. Bayesian model calibration and damage detection for a digital twin of a bridge demonstrator. Eng Rep 2023;5(11):e12669.
[19] Febrianto E, Butler L, Girolami M, Cirak F. Digital twinning of self-sensing structures using the statistical finite element method. Data-Cent Eng 2022;3:e31.
[20] Kandasamy K, Schneider J, Póczos B. Query efficient posterior estimation in scientific experiments via Bayesian active learning. Artificial Intelligence 2017;243:45–56.
[21] Wang H, Li J. Adaptive Gaussian process approximation for Bayesian inference with expensive likelihood functions. Neural Comput 2018;30(11):3072–94.
[22] Acerbi L. Variational Bayesian Monte Carlo. Adv Neural Inf Process Syst 2018;31.
[23] Acerbi L. Variational Bayesian Monte Carlo with noisy likelihoods. Adv Neural Inf Process Syst 2020;33:8211–22.
[24] Kennedy MC, O'Hagan A. Bayesian calibration of computer models. J R Stat Soc Ser B Stat Methodol 2001;63(3):425–64.
[25] Brynjarsdóttir J, O'Hagan A. Learning about physical parameters: The importance of model discrepancy. Inverse Problems 2014;30(11):114007.
[26] Kosikova AM, Sedehi O, Papadimitriou C, Katafygiotis LS. Bayesian structural identification using Gaussian process discrepancy models. Comput Methods Appl Mech Engrg 2023;417:116357.
[27] Koune I, Rózsás Á, Slobbe A, Cicirello A. Bayesian system identification for structures considering spatial and temporal correlation. Data-Cent Eng 2023;4:e22.
[28] Huggins B, Li C, Tobaben M, Aarnos MJ, Acerbi L. PyVBMC: Efficient Bayesian inference in python. 2023, arXiv preprint arXiv:2303.09519.
[29] McElreath R. Statistical rethinking: A Bayesian course with examples in R and stan. Chapman and Hall/CRC; 2018.

[30] Williams CK, Rasmussen CE. Gaussian processes for machine learning, vol. 2, MIT press Cambridge, MA; 2006.
[31] Speagle JS. A conceptual introduction to Markov chain Monte Carlo methods. 2019, arXiv preprint arXiv:1909.12313.
[32] Jeffreys H. The theory of probability. Oxford classic texts in the physical sciences, OUP Oxford; 1998.
[33] Hershey JR, Olsen PA. Approximating the Kullback Leibler divergence between Gaussian mixture models. In: 2007 IEEE international conference on acoustics, speech and signal processing-ICASSP'07. vol. 4, IEEE; 2007, p. IV–317.
[34] Rózsás Á, Slobbe A, Martini G, Jansen R. Structural and load parameter estimation of a real-world reinforced concrete slab bridge using measurements and Bayesian statistics. Struct Concr 2022;23(6):3569–600.
[35] Wosniok A, Jansen R, Chen L, Toet P, Doppenberg E, De Jong W, Chruscicki S. Static load monitoring of a concrete bridge using a high-precision distributed fiber optic sensor system. In: SMAR 2019-proceedings. 2019, p. 1–8.
[36] Martinez Colan A. Bayesian system identification of civil engineering structures using high resolution optic fibre measurements and surrogate modelling [Msc thesis], Delft, Netherlands: Delft University of Technology; 2022.
[37] Vrouwenvelder T, et al. The JCSS probabilistic model code. Struct Saf 1997;19(3):245–51.
[38] Goodman J, Weare J. Ensemble samplers with affine invariance. Commun Appl Math Comput Sci 2010;5(1):65–80.
[39] Foreman-Mackey D, Hogg DW, Lang D, Goodman J. emcee: the MCMC hammer. Publ Astron Soc Pac 2013;125(925):306.
[40] Hoffman MD, Gelman A. The No-U-Turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo.. J Mach Learn Res 2014;15(1):1593–623.
[41] Vereecken E, Slobbe A, Rózsás Á, Botte W, Lombaert G, Caspeele R. Efficient Bayesian model selection and calibration using field data for a reinforced concrete slab bridge. Struct Infrastruct Eng 2024;20(5):741–59.
[42] Ghosal S, Van Der Vaart A. Convergence rates of posterior distributions for noniid observations. Ann Statist 2007;35:192–223.
[43] Igea F, Cicirello A. Cyclical variational Bayes Monte Carlo for efficient multi-modal posterior distributions evaluation. Mech Syst Signal Process 2023;186:109868.