

Bayesian History Matching for Forward Model-Driven Structural Health Monitoring

P. Gardner, C. Lord & R. J. Barthorpe

Dynamics Research Group

Department of Mechanical Engineering, University of Sheffield
Mappin Street, Sheffield S1 3JD, UK

Abstract

Computer models are widely utilised in many structural dynamics applications, however their use depends on calibration to observational data. A complexity in calibrating a computer model is that even when the true input parameters to the model are known, there may be model discrepancy caused by the simplification or absence of certain physics. As a consequence the calibration technique employed must incorporate a mechanism for dealing with model discrepancy. Bayesian history matching is a process of using observed data in order to identify and discard areas of the computer model's parameter space that will result in outputs that are unlikely given the observational data. This is performed using an implausibility metric that encompasses uncertainties associated with observational measurements and model discrepancy. The method employs this metric to identify a non-implausible space (i.e. parameter combinations that are likely to have produced the observed outputs). A maximum *a posteriori* (MAP) approach can be used to obtain the calibrated computer model outputs from the non-implausible space. Model discrepancy between the calibrated computer model and observational data can then be inferred using a Gaussian process (GP) regression model. This paper applies Bayesian history matching in order to calibrate a computer model for forward model-driven structural health monitoring (SHM). Quantitative metrics are used to compare experimental and predictive damage features from the combined Bayesian history matching and GP approach.

Key words: History Matching, Calibration, Model Discrepancy, Emulator, Gaussian Processes

1 Introduction

Forward-model driven structural health monitoring (SHM) is a framework whereby a validated computer model is used, in a forward manner, to simulate damaged state data [1]. The computer model predictions are used to train machine learning techniques which, when shown real world observations, are capable of classifying the new data into categories such as the existence of damage, location, type and extent [2]. The success of a forward model-driven framework relies upon a computer model that is statistically representative of real world observations. Consequently, calibration and validation of this computer model is vital for producing robust decisions about structural health.

All computer models contain simplifications or the absence of certain physical processes; summarised by Box in 'All models are wrong, but some are useful' [3]. In order to account for missing physics in the calibration process, model discrepancy must be considered [4, 5]. A forward model-driven approach to SHM therefore requires a method in which model discrepancy is accounted for, producing predictions from the computer model that are statistically representative.

Bayesian history matching is a calibration methodology that reduces the computer model input space whilst accounting for model discrepancy [6–10]. The method can be used to reduce the non-implausible parameter space and produce samples of the computer models predictive distribution. A maximum *a posteriori* (MAP) estimate of the predictive distribution from the Bayesian history matching process provides an approximation of the 'most probable' calibrated computer model output given the uncertainties in the system. This MAP estimate of the output can be combined

with a Gaussian process (GP) regression model in order to infer the functional form of the model discrepancy [1]. The combined output provides an approximation of the full predictive distribution whilst considering the systems uncertainties and accounting for model discrepancy.

The focus of this paper is the application of Bayesian history matching in order to calibrate a five storey building structure under several pseudo-damage scenarios. The contribution of this paper is the use Bayesian history matching and a GP regression model to calibrate and infer model discrepancy in a finite element (FE) model. This approach allows the estimation of an approximate complete predictive distribution for the computer model whilst accounting for the uncertainties and model discrepancy in the system. Finally, the approximate predictive distribution from the approach is validated against the experimental distributions. The paper outline is as follows: Section 2 describes the Bayesian history matching methodology used in the paper with details on the calibration and inference of the model discrepancy using a Gaussian process emulator. Section 3 presents the experimental case study, the problem setup as well as demonstration and validation of results. Finally, a discussion and conclusions are presented in Section 4 and areas of further research are highlighted.

2 Bayesian History Matching

History matching is a term that originates from the oil industry and describes methods that adjust input parameters of computer models (herein referred to as *simulators*) in order to closely match data from historical production of a reservoir. Many of these approaches as reviewed by Oliver and Chen [11] are similar to the model updating methods used in SHM [12]. Craig et al. adapted the history matching approaches, outlining a Bayesian methodology that searched for all, rather than a single parameter match [6]. Bayesian history matching has been developed and applied to a variety of applications from understanding Galaxy formation [7, 8] to complex social models of HIV transfer in populations [9, 10]. The strength of this approach is that, unlike a likelihood based method, inputs and outputs from the model can be removed and added in each iteration without invalidating the analysis. Additionally, the ability to reduce the non-imausible parameter space (parameters which produce outputs similar to the experimental data) means that it is often a useful pre-calibration tool for likelihood based techniques such as Markov chain Monte Carlo based approaches. Bayesian history matching uses an implausibility metric, assessing whether a given parameter combination was not likely to have produced the output, in order to remove parts of the input space. Implausibility is a measure of distance from the experimental data to the simulator data over the processes uncertainties. This allows the creation of a statistical model form as shown in Equation 1 [4].

$$\mathbf{z}_j(\mathbf{x}) = \eta_j(\mathbf{x}, \boldsymbol{\theta}) + \delta_j + e_j \quad (1)$$

Where $\mathbf{z}_j(\mathbf{x})$ is the j th experimental output given inputs \mathbf{x} , $\eta_j(\mathbf{x}, \boldsymbol{\theta})$ is the j th simulator given \mathbf{x} and parameters $\boldsymbol{\theta}$. The model discrepancy is δ and e is the observational uncertainty. Equation 1 assumes that $\eta(\mathbf{x}, \boldsymbol{\theta})$, δ and e are independent. In order to speed up computation, Bayesian history matching utilises emulators - a statistical model of a simulator [13]. Here, a GP regression model is used as an emulator as shown in Equation 2. A GP emulator is an interpolation method that fits the simulator data exactly and also informs about the introduced code uncertainty when interpolating, due to its Bayesian formulation. A Gaussian process is a distribution over the function space and therefore allows inference to take place directly in the space of functions. This means it can be thought of as a non-parametric statistical model.

$$\eta_j(\mathbf{x}, \boldsymbol{\theta}) \sim \mathcal{GP}_j(m(\mathbf{x}, \boldsymbol{\theta}), c[(\mathbf{x}, \boldsymbol{\theta}), (\mathbf{x}', \boldsymbol{\theta}')]]) \quad (2)$$

Where \mathcal{GP} is a GP with mean function $m(\cdot)$ and covariance function $c(\cdot, \cdot)$. In the interest of conciseness the reader is referred to [13–16] for information on GPs and their use as emulators. The posterior mean of the emulator prediction, $\mathbb{E}^*[\mathcal{GP}(\mathbf{x}, \boldsymbol{\theta})]$ can be used in order to evaluate the distance between experimental and simulator data. The posterior variance of the emulator, $V_c(\mathbf{x}, \boldsymbol{\theta})$, considered as code uncertainty, is also incorporated within the implausibility measure. As a result the substitution of an emulator will not eliminate parts of the parameter space due to poor emulator predictions, until the code uncertainty is reduced. GP emulators are single output and therefore must be constructed for each simulator output. Each emulator is validated using diagnostics metrics proposed by Bastos and O'Hagan in [16]. For a single output the implausibility metric is assessed using Equation 3.

$$I_j(\mathbf{x}, \boldsymbol{\theta}) = \frac{|\mathbf{z}_j(\mathbf{x}) - \mathbb{E}^* [\mathcal{GP}_j(\mathbf{x}, \boldsymbol{\theta})]|}{[V_{o,j} + V_{m,j} + V_{c,j}(\mathbf{x}, \boldsymbol{\theta})]^{1/2}} \quad (3)$$

Where, V_o , V_m and $V_c(\mathbf{x}, \boldsymbol{\theta})$ are the variances associated with the observational, model discrepancy and code uncertainties. There are two extensions of this metric for multiple output scenarios. The first is maximum implausibility and the second is a multivariate form, Equations 4 and 5 respectively.

$$I_{max}(\mathbf{x}, \boldsymbol{\theta}) = \arg \max_j I_j(\mathbf{x}, \boldsymbol{\theta}) \quad (4)$$

$$I(\mathbf{x}, \boldsymbol{\theta}) = (\mathbf{z}(\mathbf{x}) - \mathbb{E}^* [\mathcal{GP}(\mathbf{x}, \boldsymbol{\theta})])^T (V_o + V_m + V_c(\mathbf{x}, \boldsymbol{\theta}))^{-1} (\mathbf{z}(\mathbf{x}) - \mathbb{E}^* [\mathcal{GP}(\mathbf{x}, \boldsymbol{\theta})]) \quad (5)$$

A large implausibility value (for all the implausibility metrics) indicates that the parameter set was very unlikely to produce an output that matched the experimental data given all the processes uncertainties. This means that by assessing the implausibility against a threshold, parts of the space can be eliminated. As stated in by Andrianakis et al. a sensible cut-off c is defined for the single and maximum implausibilities by Pukelsheim's 3σ rule [9] - any continuous unimodal distribution contains 99.73% of probability mass within three standard deviations from its mean [17]. For the multivariate case this can be set as a high percentile from a chi-squared distribution with j degrees of freedom [9].

After assessing the implausibility, any remaining non-implausible space is sampled from in order to propose a new input and parameter space design; the simulator is then evaluated at these points. In this paper the parameter space design is a generalised maximin Latin hypercube (GMLHC) [18], this method ensures that the design produces emulator predictions with lower predictive variance at the edges of the parameter space. Bayesian history matching can be run iteratively in waves until one of two stopping criteria are met. If all the space is identified as implausible, the model discrepancy can be increased in order to identify whether the simulator form is inadequate, leading to better model selection. The second criteria is when the emulator variance is smaller than the remaining (observational and model discrepancy) uncertainties in the system. Algorithm 1 outlines the approach used in this paper.

Algorithm 1 Bayesian History Matching for Wave k

```

 $\boldsymbol{\theta}^k \sim \text{GMLHC}$                                      ▷ Draw parameters from GMLHC
 $\mathbf{y}^k = \eta(\mathbf{x}, \boldsymbol{\theta}^k)$                          ▷ Run the simulator at parameters
Draw  $n$  samples  $\boldsymbol{\theta}_s^k \sim \mathcal{U}(\min(\boldsymbol{\theta}^k) \max(\boldsymbol{\theta}^k))$           ▷ Sample parameter space
for  $j = 1$  : no. of outputs do
    Train and validate  $\mathcal{GP}_j(\mathbf{x}, \boldsymbol{\theta}^k)$            ▷ Train and validate emulators
     $[\mathbb{E}^*[\mathcal{GP}_j(\mathbf{x}, \boldsymbol{\theta}_s^k)], V_{c,j}(\mathbf{x}, \boldsymbol{\theta}_s^k)] = \mathcal{GP}_j(\mathbf{x}, \boldsymbol{\theta}_s^k)$       ▷ Emulator predictions at  $n$  samples of  $\boldsymbol{\theta}^k$ 
    Calculate  $I_j(\mathbf{x}, \boldsymbol{\theta}_s^k)$                            ▷ Assess implausibility of samples
end for
Calculate  $I_{max}(\mathbf{x}, \boldsymbol{\theta}_s^k)$ 
for  $m = 1 : n$  do
    if  $I_{max}(\mathbf{x}, \boldsymbol{\theta}_{s,m}^k) < c$  then
         $\boldsymbol{\theta}_{nI}^k = \boldsymbol{\theta}_{s,m}^k$                                 ▷ Keep non-implausible samples
    end if
end for
bounds =  $[\min(\boldsymbol{\theta}_{nI}^k), \max(\boldsymbol{\theta}_{nI}^k)]$           ▷ Obtain new GMLHC bounds
if any  $(V_{c,j}^k(\mathbf{x}, \boldsymbol{\theta}) < (V_{o,j} + V_{m,j}))$  or isempty( $\boldsymbol{\theta}_{nI}^k$ ) then
    Stop                                              ▷ Stop if stopping criteria are met
end if

```

2.1 Calibrated Simulator Prediction and Model Discrepancy Estimation

Bayesian history matching produces a set of non-implausible samples; contained within this set will be the posterior of the parameters given the data, $p(\boldsymbol{\theta}|\mathbf{z})$. The non-implausible samples from the final wave can be used as an

approximation of this posterior. The resulting emulator outputs for the final non-implausible set are estimates of the predictive simulator distribution, given the uncertainties described. A MAP estimate of the predictive distribution for each output will be adopted as the ‘most probable’ calibrated simulator.

The calibrated simulator will have the limitation that even given the calibrated parameter set the simulator will not produce outputs that fully describe the experimental data due to model discrepancy. In order to infer the functional form of the model discrepancy a GP regression model (with noise) can be fitted to the MAP estimate of the predictive distribution and the same training experimental data points used in the Bayesian history matching process. The resulting prediction provides a Gaussian approximation of the predictive distribution of the simulator accounting for the assumed uncertainties.

3 Five Storey Building Structure

Predictions of modal frequencies for different damage extents were performed on a five storey building structure. The focus of this paper is demonstrating that the distributions of these predictions are representative of those obtained experimentally: the most important consideration in forward model-driven SHM. Modal testing of a representative five storey building structure, made from aluminium 6082, was performed under different pseudo-damage extents, in this case added masses. The structure (Figure 1a) had masses in increments of 0.1kg, from 0kg to 0.5kg, fixed to the first floor, (Figure 1b). The structure was excited with Gaussian noise with a 409.6Hz bandwidth via an electrodynamic shaker. Sampling rate and sample time were chosen to allow a frequency resolution of 0.05Hz. Accelerometers were placed at each of the five floors in order to obtain the first five bending modes. 40 averages were acquired for each measurement and ten repeats were performed for each damage extent, in order to estimate the distribution of the modal frequencies. Feature selection found that the second and third bending natural frequencies, ω_2 and ω_3 , were most sensitive to the presence of damage. As a consequence these features were the target outputs of the simulator.

The simulator, a modal FE model, was run for the six damage extents. The second and third bending natural frequencies were extracted from the simulator as the outputs \mathbf{y} . Prior bounds on the parameters, $\boldsymbol{\theta}$, shown in Table 1, were $\pm 10\%$ of typical material properties for aluminium 6082. The inputs were added mass in kg, $\mathbf{x} = \{0, 0.1, \dots, 0.5\}$.

3.1 Calibration via Bayesian History Matching

Calibration of the simulator was performed using Bayesian history matching as outlined in Algorithm 1. The experimental data $\mathbf{z}(\mathbf{x}_z)$ used in the Bayesian history matching process were the mean natural frequencies when $\mathbf{x}_z = \{0, 0.3, 0.5\}$ kg. The unseen validation set were the full repeat measurements when $\mathbf{x}_z = \{0.1, 0.2, 0.4\}$ kg. This highlights that with a small subset of damage data predictions can be made using forward model-driven SHM.

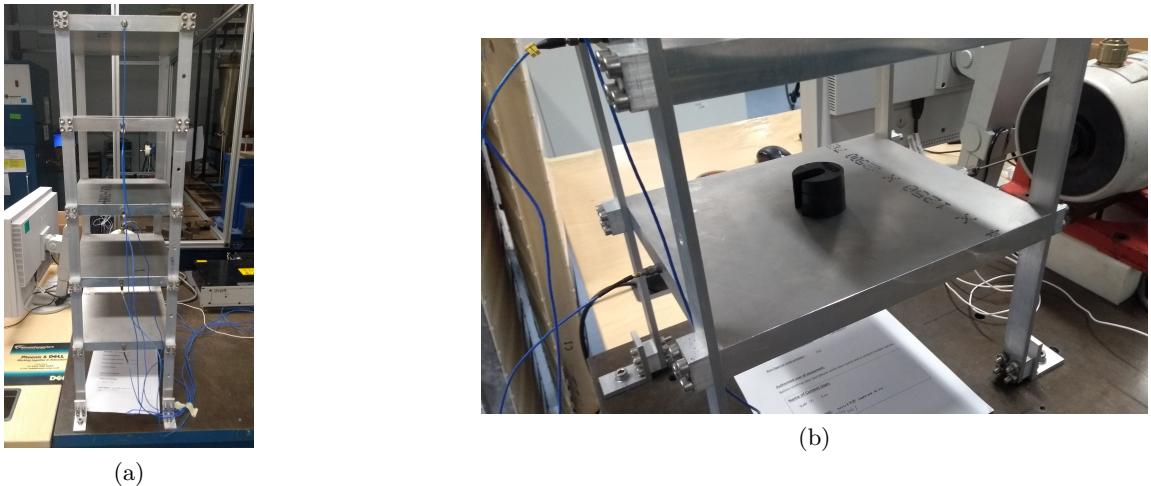


Figure 1: (a) Five storey building structure setup (b) Pseudo-damage - added masses.

Parameter	Lower Bound	Upper Bound
Elastic Modulus, E , GPa	63.9	78.1
Possions Ratio, ν	0.297	0.363
Density, ρ , kg/m ³	2493	3047

Table 1: Calibration parameter prior bounds.

Two separate 50 and 10 point GMLHC designs (one for training the emulators and one for validation) were obtained in order to fill the initial parameter space. The simulator was run with these parameters combinations for the six inputs, \mathbf{x} . The simulator outputs were used to train and validate two emulators for ω_2 and ω_3 . Constant mean and a squared exponential covariance functions were applied in the emulator due to the smooth functional form of the simulator. A small nugget of $v = 1 \times 10^{-6}$ was implemented in order to improve the numerical stability of the emulator whilst minimally affecting the solution as shown by Andrianakis and Challenor [19]. The maximum implausibility metric with a 3σ threshold was implemented. A sample size of 500,000, drawn from a 3 dimensional uniform distribution, were used in order to assess the parameter space.

The first wave calculated a non-implausible space $\approx 6.5\%$ of the original space, highlighting the method's strength in removing unlikely parameters from the analysis. After the first wave the second stopping criteria was met (code uncertainty was less than the remaining uncertainties). This means that the functional form of the emulator was well defined from the GMLHC parameter space and accurately interpolated across the complete parameter space. In order to visualise the non-implausible space, minimum implausibility and optical depth plots were created. These plots divided the parameter space into bins in which the samples are placed. Minimum implausibility is the lowest implausibility found in the bin for each parameter combination. If the number of samples is large enough, these plot provide an estimate of what parts of the input space can be discarded irrespective of the other parameters. The optical depth plots are constructed in a similar method to the minimum plausibility but instead show an estimate of the probability of a point being non-implausible given the bin's set of parameters. This is the ratio of the number of non-implausible points divided by the number of samples in the bin. Figure 2 demonstrates these plots for the first wave. Here it can be seen that high values of elastic modulus and low values of density are identified as non-implausible with Possion's ratio being relatively insensitive to the outputs. There is a clear linear correlation

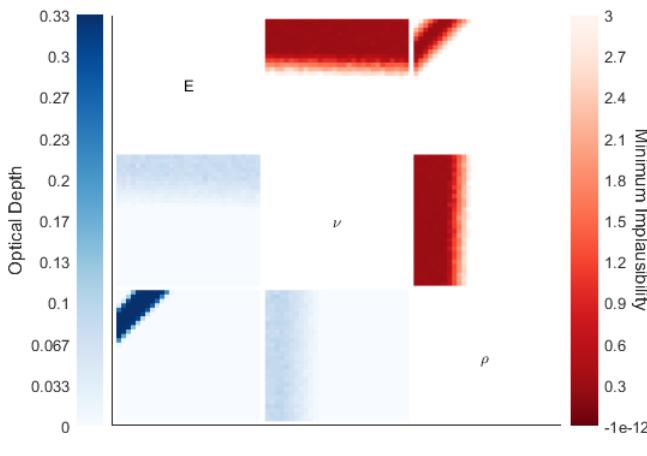


Figure 2: Minimum implausibility and optical depth plots for the first wave of Bayesian history matching. Each quadrant is a comparison of two parameter combinations for the given metric, e.g. the top right and bottom left quadrant are ρ against E for minimum implausibility and optical depth respectively.

Uncertainty	ω_2	ω_3
Observational		
V_o ,	0.02	0.08
Model Discrepancy		
V_m	0.01	0.01

Table 2: Uncertainties in implausibility measure.

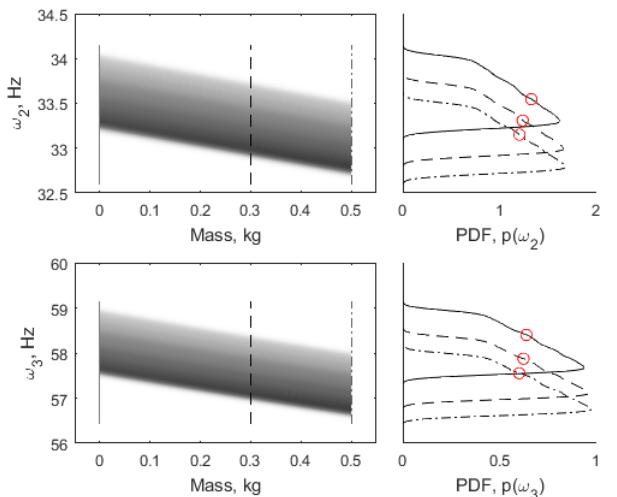


Figure 3: Density of outputs from non-implausible samples (left) and kernel density estimates (KDEs) of the predictive distribution (right); o are experimental data points. Slices of the densities at the training experimental inputs \mathbf{x}_z are shown for the 0kg (-), 0.3kg (---), and 0.5kg (--) masses.

between the non-imausible space of the elastic modulus and density, displayed in the bottom left and top right quadrants of Figure 2.

The non-imausible samples from the first wave can be used to approximate the predictive densities for simulator outputs, presented in Figure 3. Here both densities are shown to be negatively skewed. The right of Figure 3 shows the predictive densities at the experimental input data points, \mathbf{x}_z where the red circles are the experimental data points $\mathbf{z}(\mathbf{x}_z)$. The experimental data points lie slightly below the distribution's mode and close to the mean, but firmly within the non-imausible distribution.

The MAP estimates of the predictive distributions for each output from Figure 3 were used as inputs for two independent GP regression models (with noise). These GPs were utilised to infer the model discrepancy and observational uncertainty for each output (ω_2 and ω_3). The modelling assumptions for both GPs were zero mean and squared exponential covariance functions; this represented the belief that no knowledge of the mean functional forms were known *a priori* but the functions were believed to be relatively smooth. Figure 4 presents the two GP outputs for each natural frequency, both showing good agreement with the experimental data.

3.2 Validation of Predictive Distributions

Validation of the predictive distributions is an important aspect of forward model-driven SHM. Ensuring that the predictive distributions of the calibrated simulator are valid for the experimental data provides confidence that the simulator outputs may be used in training a classifier to perform SHM. In order to validate the predictive distributions, normalised mean squared errors and KL-divergences have been calculated.

The normalised mean squared errors (NMSEs) for each output compared to the experimental data were 0.04 and 0.02. This demonstrates the very good agreement between the experimental data and the predictive distribution's means. The mean percentage difference in the variances for each output were 0.2% and 0.8% demonstrating that for the lower order statistical modes the predictive distribution is in good agreement with the experimental data.

To validate the performance of both a Gaussian approximation of the experimental data and the predictive distributions for each output, the Kullback-Leibler (KL) divergence was employed. KL-divergence is a measure of relative entropy and can be defined as the average number of extra bits (binary digits) required to encode the data given that distribution q is used to model the ‘true’ distribution p . A KL-divergence equal to zero implies $q = p$. In both cases the KDE of the experimental data was taken to be the ‘true’ distribution. Figure 5 presents the KL-divergence

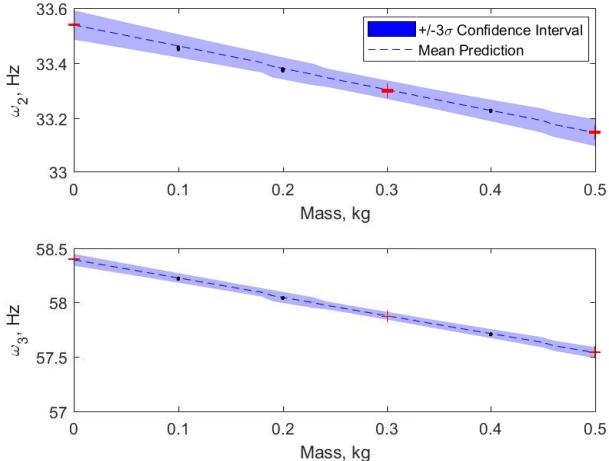


Figure 4: Predictive distributions from the combined Bayesian history matching and GP approach, + are experimental training points, · are experimental test points.

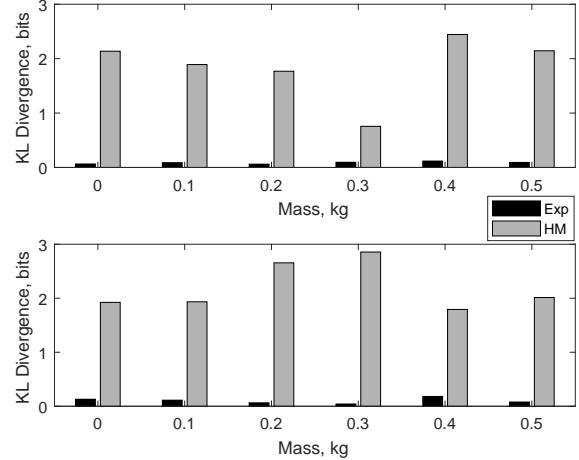


Figure 5: KL-divergence compared to the KDEs from the experimental data. Exp denotes Gaussian approximation of experimental data (black); HM denotes the predictive distribution of the combined Bayesian history matching and GP approach (grey).

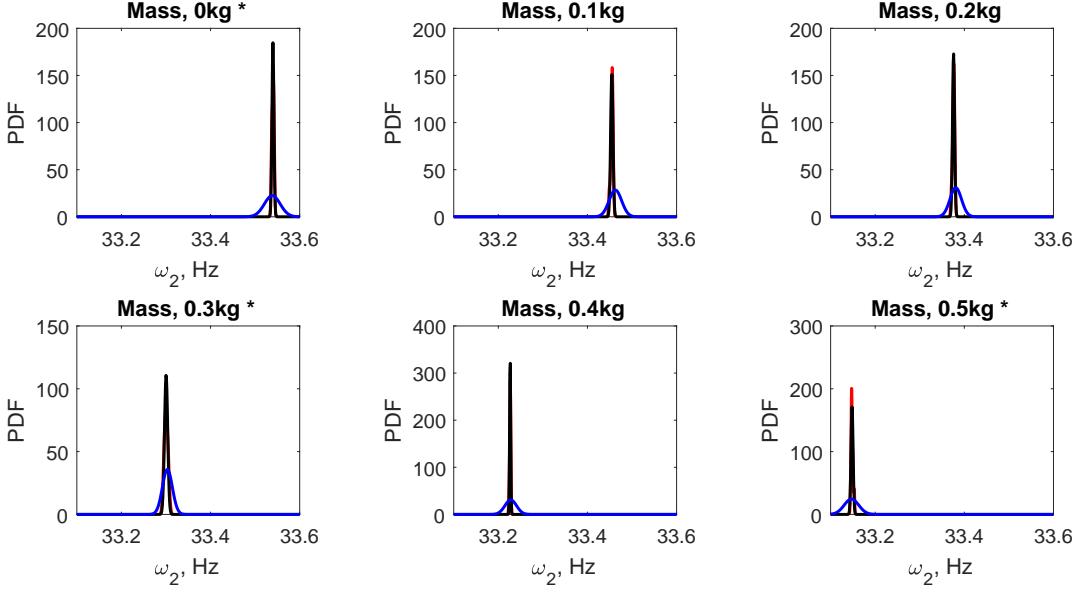


Figure 6: Predictive distributions compared to experimental data for ω_2 ; KDE from experimental data (red), Gaussian approximation of experimental data (black), predictive distribution from Bayesian history matching and GP approach (blue); * denotes inputs used as training data.

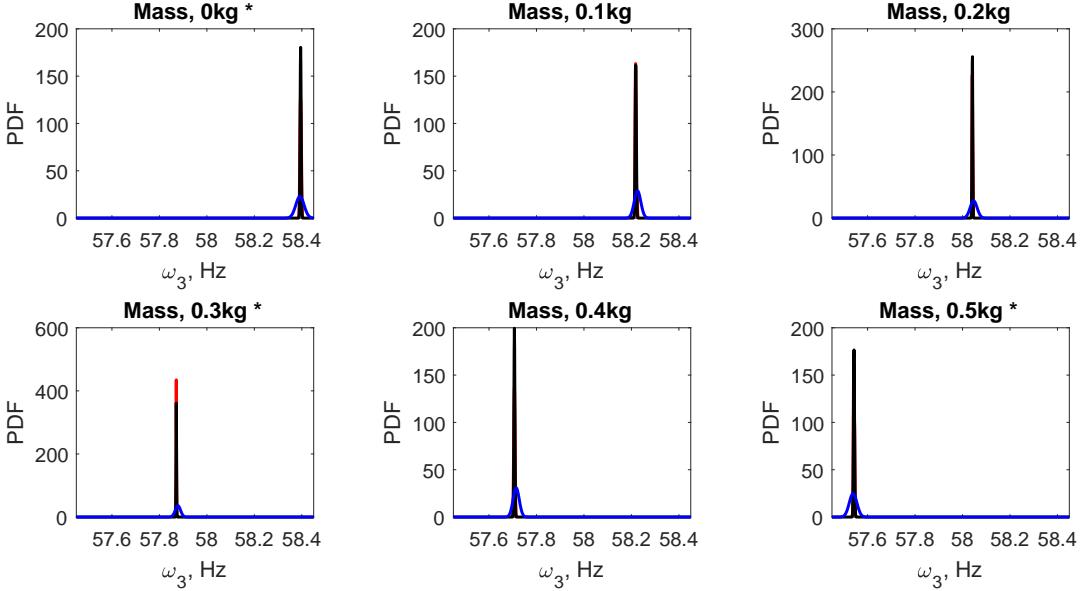


Figure 7: Predictive distributions compared to experimental data for ω_3 ; KDE from experimental data (red), Gaussian approximation of experimental data (black), predictive distribution from Bayesian history matching and GP approach (blue); * denotes inputs used as training data.

for each output (top ω_2 , bottom ω_3). It can be seen that a Gaussian approximation of the experimental data is appropriate for these two features, indicated by near zero KL-divergence. Additionally, the KL-divergence for the predictive distributions for each output are below 3 and most are close to 2. This means that very few additional bits would be needed in order to encode the kernel density estimate (KDE) fully. It is noted that the performance of the KL-divergence is similar throughout the damage extents demonstrating good predictive performance away from the training experimental data. The summed KL-divergence for the Gaussian approximations of the experimental data

for each output were 0.5 and 0.6 whereas the predictive distributions were 11.1 and 13.0. Finally, a visual comparison in Figures 6 and 7 of the distributions for each damage extent confirms good mean predictions and slightly large variance estimates in the predictive distributions compared to the experimental data.

4 Discussion and Conclusions

Predictions of damage extents for two natural frequencies of a representative five storey building structure were performed using a combined Bayesian history matching and GP regression model approach. One wave of Bayesian history matching was required in order to find a non-imausible parameter space. Samples from this space generated an approximate predictive distribution for each output from which a MAP estimate was taken. Two GP models were built in order to infer the model discrepancy and observational uncertainties. The predictive distributions from the combined approach for each output were validated and demonstrated to be representative of the experimental data, as shown in low NMSEs and low KL-divergences, on and unseen test set. This highlights the approaches effectiveness as a method for forward model-driven SHM.

It is noted Bayesian history matching is likely to be of greater benefit in more complex cases, especially when less is known about the parameter space and the quality of the simulator. The method also has similar difficulties to conventional calibration techniques as the functional form of the emulator restricts the non-imausible space, unless the model discrepancy variance is increased. Bayesian calibration with bias correction, another calibration technique that considers model discrepancy, has the strength that the method allows more flexibility in inferring calibrated parameters whilst simultaneously estimating the functional form of the model discrepancy [4, 20].

Areas for further research are predominately associated with improving Bayesian history matching. To speed up the algorithm computationally, sparse GP emulators could be implemented [21]. Additionally, better methods of sampling the non-imausible space such as a maximum entropy approach could be used. This would mean that new simulator runs are only implemented at points that provide the most information gain. A method for accurately estimating the posterior of the parameters $p(\theta|z)$ from the non-imausible space should be investigated. This would allow sensitivity analysis and a full Bayesian treatment of the outputs to be conducted, rather than a MAP estimate.

References

- [1] P. Gardner, R. J. Barthorpe, and C. Lord, “The Development of a Damage Model for the use in Machine Learning Driven SHM and Comparison with Conventional SHM Methods,” in *Proceedings of ISMA2016 International Conference on Noise and Vibration Engineering*, 2016, pp. 3333–3346.
- [2] A. Rytter, “Vibrational Based Inspection of Civil Engineering Structures,” Ph.D. dissertation, 1993.
- [3] G. E. P. Box and N. R. Draper, *Empirical Model-Building and Response Surfaces*, 1987.
- [4] M. C. Kennedy and A. O’Hagan, “Bayesian calibration of computer models,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 63, no. 3, pp. 425–464, 2001.
- [5] J. Brynjarsdóttir and A. OHagan, “Learning about physical parameters: the importance of model discrepancy,” *Inverse Problems*, vol. 30, no. 11, p. 114007, 2014.
- [6] P. S. Craig, M. Goldstein, A. H. Seheult, and J. A. Smith, “Pressure Matching for Hydrocarbon Reservoirs: A Case Study in the Use of Bayes Linear Strategies for Large Computer Experiments,” in *Lecture Notes in Statistics*, 1997, pp. 37–93.
- [7] I. Vernon, M. Goldstein, and R. G. Bower, “Galaxy formation: a Bayesian uncertainty analysis,” *Bayesian Analysis*, vol. 5, no. 4, pp. 619–669, dec 2010.
- [8] I. Vernon, M. Goldstein, and R. Bower, “Galaxy Formation: Bayesian History Matching for the Observable Universe,” *Statistical Science*, vol. 29, no. 1, pp. 81–90, feb 2014.

- [9] I. Andrianakis, I. R. Vernon, N. McCreesh, T. J. McKinley, J. E. Oakley, R. N. Nsubuga, M. Goldstein, and R. G. White, “Bayesian History Matching of Complex Infectious Disease Models Using Emulation: A Tutorial and a Case Study on HIV in Uganda,” *PLoS Computational Biology*, vol. 11, no. 1, 2015.
- [10] I. Andrianakis, I. Vernon, N. McCreesh, T. J. McKinley, J. E. Oakley, R. N. Nsubuga, M. Goldstein, and R. G. White, “History matching of a complex epidemiological model of human immunodeficiency virus transmission by using variance emulation,” *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, vol. 66, no. 4, pp. 717–740, aug 2017.
- [11] D. S. Oliver and Y. Chen, “Recent progress on reservoir history matching: a review,” *Computational Geosciences*, vol. 15, no. 1, pp. 185–221, jan 2011.
- [12] M. I. Friswell, “Damage identification using inverse methods,” *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 365, no. 1851, pp. 393–410, feb 2007.
- [13] A. O’Hagan, “Bayesian analysis of computer code outputs: A tutorial,” *Reliability Engineering and System Safety*, vol. 91, no. 10-11, pp. 1290–1300, 2006.
- [14] A. O’Hagan and J. Kingman, “Curve Fitting and Optimal Design for Prediction,” *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 40, no. 1, pp. 1–42, 1978.
- [15] C. E. Rasmussen and C. K. I. Williams, *Gaussian processes for machine learning.*, 2004, vol. 14, no. 2.
- [16] L. S. Bastos and A. O’Hagan, “Diagnostics for Gaussian Process Emulators,” *Technometrics*, vol. 51, no. 4, pp. 425–438, 2009.
- [17] Pukelsheim, “The three sigma rule,” *American Statistician*, vol. 48, no. 2, pp. 88–91, 1994.
- [18] H. Dette and A. Pepelyshev, “Generalized Latin Hypercube Design for Computer Experiments,” *Technometrics*, vol. 52, no. 4, pp. 421–429, nov 2010.
- [19] I. Andrianakis and P. G. Challenor, “The effect of the nugget on Gaussian process emulators of computer models,” *Computational Statistics & Data Analysis*, vol. 56, no. 12, pp. 4215–4228, dec 2012.
- [20] P. Gardner, R. J. Bartherope, and C. Lord, “Bayesian Calibration and Bias Correction for Forward Model-Driven SHM,” in *Proceedings of IWSHM2017 International Workshop of Structural Health Monitoring*, 2017, pp. 2019–2027.
- [21] M. Titsias, “Variational Learning of Inducing Variables in Sparse Gaussian Processes,” *Aistats*, vol. 5, pp. 567–574, 2009.