Stochastic model updating for parameter identification using Bayesian inference

P. Pandey [1], **H. Haddad. Khodaparast** [1], **M. I. Friswell** [1], **T. Chatterjee** [1], **T. Deighan** [2]
[1] Swansea University, Faculty of Science and Engineering, Department of Aerospace Engineering, Swansea, Wales, United Kingdom

[2] United Kingdom Atomic Energy Authority
United Kingdom

Abstract
Modelling the uncertainties at the joint interface is crucial as it affects the dynamics of the whole system. But modelling the interface using only a physics-based (mathematical) model may lead to large differences between predictions and observations as the physics of the actual system is only partially known. The research focuses on the Bayesian stochastic model updating of joints in physical structures using Bayesian inference. Substructures (joint structures) are modelled using physics-based models. In this work, different non-linear models are used to generate dynamic responses. Using the dynamic responses, the backbone curve is generated which is used to establish the likelihood function for Bayesian identification of the joint parameters. The backbone curves are the measurement data (with added noise) which is used by the Metropolis-Hastings algorithm in Markov-Chain Monte Carlo (MCMC) sampling. The algorithm is applied for different non-linear cases for parameter identification using Bayesian inference

1 Introduction

Joints in the structure transmit loads and moments between different substructures inducing nonlinearity in the system, changing the overall damping and stiffness of the structure thus resulting in a change in the dynamic response and inadequate damage characterisation. Quantifying the parameter accounting for the uncertainties in the contact interface can be a bit tedious. These uncertainties can be due to various factors such as geometric, material non-linearities at the interface resulting from friction, roughness, or different geometric and material properties of the interacting parts [1]. Due to the uncertainties, non-linear dependency of the dynamic response with stiffness and damping of the structure was observed [2]. This hinders the modelling of the physics at the interface thus affecting the prediction of system response under regular excitation. Modelling the uncertainties and dynamics at the joint interface using a mathematical model may lead to large differences between predictions and observations. This can be due to various errors induced in the model due to uncertainties in the governing equations, boundary conditions etc. [3]. These errors can be minimized through model updating which uses experimental data of the structure to ascertain its unknown parameters. Categorised as deterministic and stochastic, deterministic model updating finds the parameter values of a specified system model such that it minimizes the difference between experiments and predictions [4]. However, in order to account for the errors (due to incorrect modelling, missing information etc), and uncertainties present in the model, stochastic model updating is adopted where the set of model parameters is generated which converges on the set of experimental results [5]. The stochastic model updating approach can be expressed in Bayesian or Frequentist way. The Bayesian approach makes use of the posterior distribution to encapsulate the information of the unknown variables whereas the frequentist approach relies on the sampling distribution taken from numerous data sets after repeated trials [6]. The current work focuses on stochastic model updating using the Bayesian approach.
Mares et. al [7] used MCMC method for model updating on 3 degrees of freedom mass-spring system using their experimental frequency response functions to update the spring stiffness parameters. The structural

models were updated using ground acceleration data from the earthquake on reinforced concrete structure utilising a simulation-based Bayesian methodology for system identification and using these updated models, reliability of the structural system was estimated [8]. Sun et.al [9] used hieratical Bayesian inference with MCMC algorithm to draw samples of stiffness parameters for a 9 storey shear-type building with incomplete modal data.
In this paper, Bayesian inference with Metropolis Hastings algorithm is used to draw samples of parameters for different non-linear cases. The novelty of the work lies in the likelihood function considered here. Backbone curves extracted from the dynamic response of the system are used for defining the likelihood function. The work is organized as follows. Section 2 briefly explains the concept of the backbone curve and how it is generated for the system with different non-linearities. Section 3 describes the parameter identification using Bayesian inference with a light on prior, likelihood and finally posterior distribution for the problem at hand. In Section 4, the theory behind the MCMC sampling is illustrated. Section 5 consists of analysis of the results of the sampling for different non-linear cases. The paper is concluded in Section 6.

2 Mathematical modelling for data generation

2.1 Backbone curve

Response of the non-linear system under forced excitation is significant but can be complicated to interpret. Backbone curves can help in understanding the non-linearity in the system without any forced excitation. It employs the solution of a non-linear system where the natural frequency is described as a function of the amplitude of the system's response with an assumption that forcing and damping are not present [10]. Furthermore, the estimated instantaneous frequency is assumed not to be rapidly altered by dissipative forces acting on the system. For the linear system, the backbone curve is a straight line whereas for the non-linear system it deviates from the vertical line (Fig. 1). Backbone establishes a base for identifying the non-linearity in the system For generating a backbone curve (for SDOF or MDOF), the response of a system (Eq. 1) is considered where F(x,$\dot{x}$) represents different non-linearities in the system.

$$m\ddot{x} + C\dot{x} + Kx + F(x, \dot{x}) = 0 \tag{1}$$

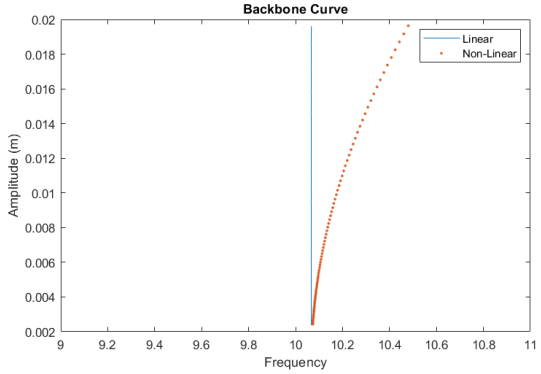


Figure 1: Backbone curve for linear and non-linear system

2.2 Resonant decay method

A free decay response with a single mode of vibration is a resonant decay response. The structure is excited at a relevant frequency using harmonic excitation and when the structure reaches the desired resonance condition, the input force is removed and the model undergoes a decayed response at a single mode of vibration [11]. The instantaneous frequency and amplitude envelope of this decayed response gives the

backbone curve of the system. Resonant decay response can be measured for SDOF and multi-degree of freedom (MDOF) systems. For an SDOF system, every free decay is resonant decay. However, for an MDOF system, harmonic excitation is given at a single frequency of the mode interested.

2.3 Extracting backbone curve using resonant decay method

For each case of non-linearity, a range (upper and lower bound) is defined for each parameter. From this defined range, different values for the parameters are generated using Latin Hypercube Sampling (LHS). Using these parameter values, system response is generated which is further used for the extraction of backbone curves. Instantaneous amplitude and frequency are obtained using the decay response for backbone curve estimation. Instantaneous frequency (f_i) and instantaneous amplitude (A_i) are calculated using maximum and minimum peak points of the response signal.

$$f_i = (t_{i+1} - t_i)^{-1} \tag{2}$$

$$A_i = \frac{1}{2}(\text{Max}(\text{Amplitude}(i)) - \text{Min}(\text{Amplitude}(i))) \tag{3}$$

where t_i and t_{i+1} are the instantaneous time corresponding to the maximum peak points (see figure 2). For a non-linear system, (represented in Eq. 1) where F represents the non-linear damping and restoring forces, the resonant decay method is used for approximating the backbone curve proposed in [12]. This decaying response is used to evaluate instantaneous frequency and amplitude to estimate the backbone curve. For a single degree of freedom (SDOF) system, to excite the system to resonance, the initial condition (either displacement or velocity) large enough to excite the non-linearity is considered.
Figure 2 shows the free decaying response of one of the cases considered (Case 1: Cubic stiffness). This decaying response is simulated experimental data from where the backbone curve is drawn using the instantaneous frequencies and amplitude from the response. This approach is used for five different non-linear cases which will be introduced later in the paper.

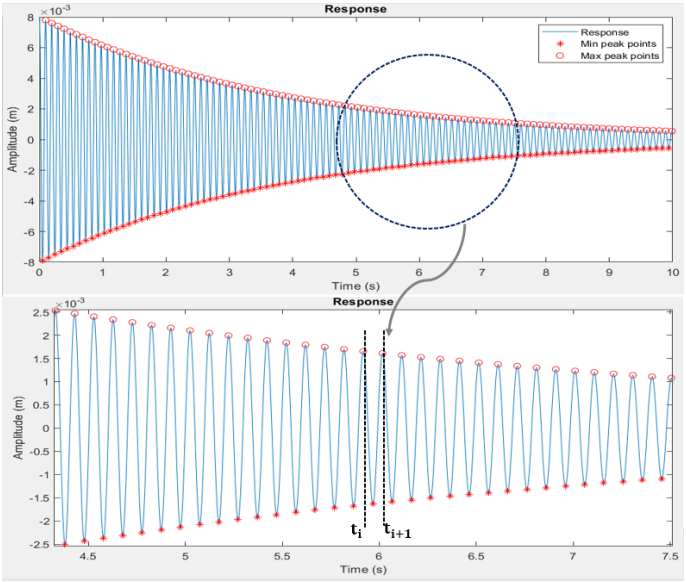


Figure 2: Response decay method with maximum and minimum peak points

In this work, SDOF with three different non-linear systems (Eq. 4, 6) is considered where an initial condition of displacement is taken to excite the non-linearity. For each non-linear case, around 100 samples of parameters are generated using LHS sampling and each set of parameters was used to generate backbone curves. The equations governing the non-linear systems which are shown below are used for generating the

decaying response for each non-linear system
Case 1: Cubic stiffness model

$$m\ddot{x} + C\dot{x} + K_1 x + K_2 x^3 = 0 \tag{4}$$

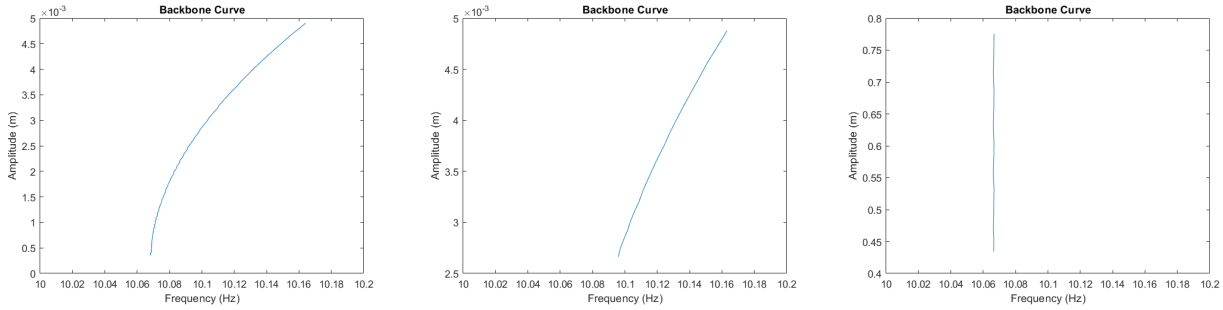
Case 2: Quadratic damping with cubic stiffness model

$$m\ddot{x} + C_1\dot{x} + C_2\dot{x}|\dot{x}| + K_1 x + K_2 x^3 = 0 \tag{5}$$

Case 2: Dry friction model

$$m\ddot{x} + C_1\dot{x} + C_2 \operatorname{sign}\dot{x} + Kx = 0 \tag{6}$$

Figure3 shows extracted backbone for different non-linear systems. These estimated backbone curves represent the possible measurements of the system. Here, the inverse problem where the parameters of interest are calculated based on the observations is performed using Bayesian inference.



(a) Cubic stiffness (b) Quadratic damping with cubic stiffness (c) Dry friction

Figure 3: Backbone curve of the cases using response decay method

3 Bayesian inference for parameter identification

The parameters to be identified are represented as random variables with a Probability distribution function. Based on previous information about the distribution and observations, Bayesian Inference creates a joint Probability distribution function for the parameters. The well-known Bayes Rule underpins the Bayesian Inference framework as:

$$P(\theta \mid D, M) = \frac{P(D \mid \theta, M) \cdot P(\theta \mid M)}{P(D \mid M)} \tag{7}$$

$P(D \mid \theta, M)$ and $P(\theta \mid M)$ are the likelihood function and prior distribution respectively. The likelihood function represents the degree of agreement between the measurement and the system model. Prior distribution reflects a priori knowledge or initial assumption about the model parameters before any measurements are made. The product of likelihood and prior can be defined as the posterior distribution $P(\theta \mid D, M)$. $P(D \mid M)$ is the normalisation constant of the posterior distribution. D is the measurements made from M models (system) to estimate parameter vector θ.

In this paper, θ are the vector of system parameters such that $\theta \in R^{mx1}$ where m is the number of parameters. $(D \mid M)$ corresponds to the set of measurements i.e. the backbone curves for the given model (M). Uniform distribution is assumed for each parameter with their corresponding lower and upper bounds and $P(\theta \mid M)$ which is prior knowledge of parameters is the joint probability density function (pdf) of the parameters.

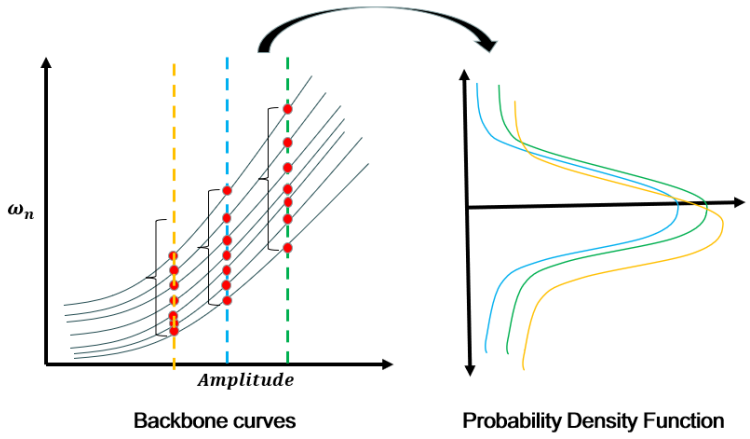


Figure 4: Likelihood function from backbone curves

For this problem, the likelihood function is defined from the points represented in Figure. 4. The red dots in the figure correspond to the points in the backbone curve at a certain amplitude. A set of such points at an amplitude gives the distribution for the likelihood function. To capture the non-linearity better, the pdfs are considered around a higher amplitude region, From the generated pdfs, it was observed that the distribution was close to the uniform distribution for all three cases. Hence the likelihood function follows the uniform distribution.

Posterior distribution combines our prior knowledge of the parameters and the updated knowledge from the measurements. To accommodate the loglikelihood, posterior is defined in log space and is expressed as $\log(P(\theta \mid D, M))$ such that $\log(P(\theta \mid D, M)) = \log(P(D \mid \theta, M)) + \log(P(D \mid M))$.

4 Theory of MCMC sampling

Markov Chain Monte Carlo (MCMC) sampling is used to generate samples from the target distribution without much information on the distribution's mathematical aspects, such as the normalisation constant $P(D)$ [13]. These samples can be used to estimate the mean of the posterior distribution where posterior distribution is the product of prior and likelihood. MCMC sampling utilizes properties of Monte Carlo and Markov Chains at the same time using the Metropolis-Hastings algorithm to decide on the acceptance of the samples. Monte Carlo defines the method of generating random numbers from the proposal distribution whereas Markov Chains are a sequence of numbers where each number is dependent on the number previous to it in the sequence. The metropolis-Hastings algorithm helps in deciding which proposed values of the parameter should be accepted or rejected. Samples are selected from the proposed distribution (Monte Carlo) with a mean equal to the previous parameter value for MCMC sampling (Markov Chains). When samples are created using MCMC, the trace plot appears to wander and is frequently referred to as a Random walk. The density plot that results in does not mirror the proposal density. The ratio between the posterior probability of the sample and the sample drawn before is calculated once a sample is generated from the proposal distribution (h).

$$h\left(\theta_{n+1}, \theta_n\right) = \frac{\text{Posterior probability of } \theta_{n+1}}{\text{Posterior probability of } \theta_n} \tag{8}$$

If the posterior probability of the sample drawn is greater than the posterior probability of the previous sample ($h > 1$), then the new sample is accepted. However, if $h < 1$, the new sample is not yet discarded instead it is treated as in acceptance probability (α) defined in the equation below.

$$\begin{aligned} \text{Acceptance probability} &= \alpha\left(\theta_{\text{n+1}}, \theta_n\right) \\ &= \min\left[h\left(\theta_{n+1}, \theta_n\right), 1\right] \end{aligned} \tag{9}$$

A random number (r) from a uniform distribution is drawn and is compared with the Acceptance probability (α). If α is greater than r, then the new sample is accepted else rejected and the previous sample is considered a new sample. The samples generated using MCMC are samples from the posterior distribution.To start the MCMC sampling, candidate samples are drawn from the proposal distribution. For the current analysis, the normal distribution is chosen as the proposal distribution. To avoid dependence of MH algorithm samples on the beginning values the first few samples are eliminated. This is known as the burn-in phase. This burn-in period varies based on the trace and is the time it takes for the chain to stabilise and avoid drifting. Furthermore, because the parameter samples are created using Markov chains, they can be highly autocorrelated even after accurately describing the model. So, to lessen the autocorrelation, thinning can be implemented which entails increasing the number of samples and obtaining samples at regular intervals. However, thinning is not implemented for this analysis. There are some literature which does not support using thinning for reducing autocorrelation as throwing away data is not a precise way of increasing the efficiency of Markov chains [14][15].

4.1 Performance diagnostics for MCMC samples

Out of several ways to diagnose the performance of MCMC samples, few of them (such as acceptance rate, trace plot, and KL divergence) have been explored in this paper to check the performance of MCMC.
The acceptance rate represents the percentage of times the MCMC sampling generates parameter samples that is different from the prior sample. An acceptance rate of 0.25 means, that while creating a sequence of parameter values, the algorithm creates new parameter values 25% of the time while staying on the old parameter values 75% of the time.
The Trace plot is drawn to observe the sample values acquired as a check for sample convergence about the mean value. A trace plot is made to observe the convergence of the simulated Markov chain to its stationary distribution. It gives the history of the parameter value across iterations of the chain.
There are two distinct probability distributions, such as the true distribution and an approximate version of it, for each parameter. True distribution is the one from which measurement data (i.e backbone cure) was taken whereas approximate distribution is derived from the MCMC samples of each parameter. The degree to which one probability distribution deviates from another is measured by the Kullback-Leibler Divergence score, or KL divergence score (KLD Score). If the score is 0, it implies that both distributions are equally likely; if not, it is positive.

5 Analyzing the samples from MCMC Sampling

The prior, likelihood and the posterior described in Section 3 is considered for MCMC sampling. The system with cubic stiffness (i.e. Case 1) is shown for preliminary analysis of the MCMC samples. A similar analysis is performed for each case but only Case 1 is shown in this work. True distribution defines how the parameters in the measurement data are distributed. Here, the joint probability density function (pdf) of the parameters is taken where each parameter is uniformly distributed with lower and upper bounds defined in Table 1. Figure 5 and 6 show the distribution of the 5000 samples (of each parameter) drawn from the posterior and their corresponding trace plot respectively.

Table 1: Range of parameters of measurement data defining true distribution for Case 1.

Parameters	Maximum	Minimum
K_1 (N/m)	6000	7000
K_2 (N/m^3)	6000000	6500000
C (Ns/m)	0.2	2

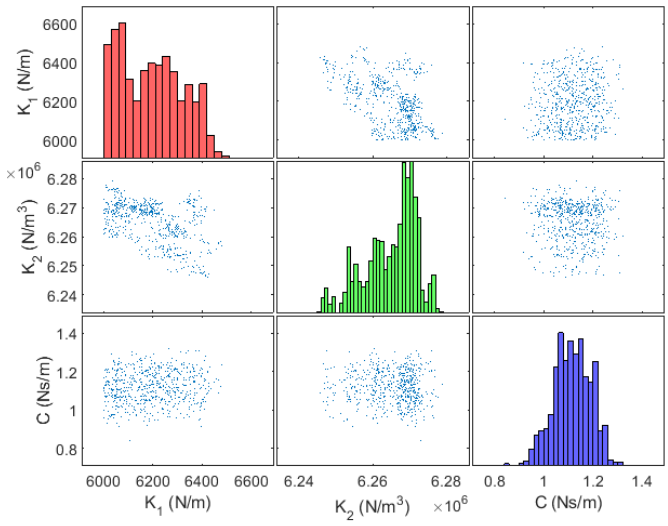


Figure 5: Scatter plot Matrix of 5000 samples of K_1 (red), K_2(green) and C (blue) for Case 1.

From the distribution of K_1 and C (Figure 5) it can be discerned that the range is more or less near to the true distribution although the same cannot be said for K_2. The trace plot for C shows that with the given proposal distribution, the algorithm performs effectively and despite the correlation in samples, the Markov chain mixes nicely and the samples approach stationary distribution faster. Also, a high correlation in the samples of K_1 and K_2 can be seen through the trace plot (Figure 6). The proposal distribution for K_1 and K_2 does not play well with the algorithm. The jump in the proposal (i.e. step size) is quite small for K_1 and K_2 (relatively small for K_2), resulting in the slow movement of the state through state space, therefore traversing the posterior distribution takes a long time. A high level of correlation was observed between the samples over periods of time i.e.autocorrelation. Autocorrelation measures the linear dependency of the current value of the chain to its past values and varies between -1 to 1. For the same number of samples, information derived from dependent samples about the stationary distribution is far less than the information derived from independent samples. From Figure. 7, where the first 100 samples are considered and it can be seen that autocorrelation for C for first 100 and last 100 samples is comparatively less than corresponding values for K_1, K_2. Also, no reduction in autocorrelation values was observed during the end of the iteration.

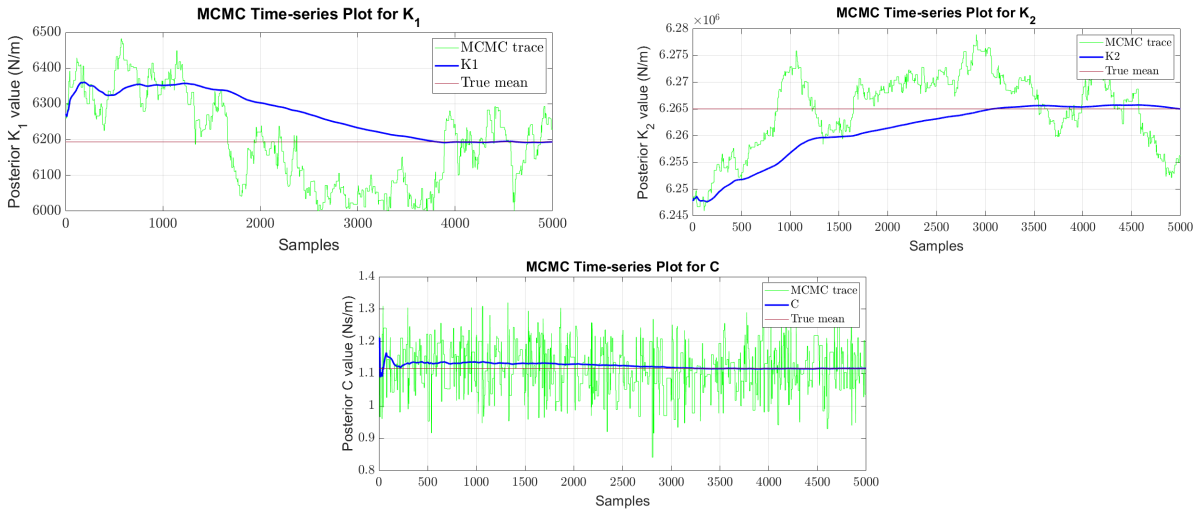


Figure 6: Trace plot for K_1, K_2 and C with 5000 MCMC samples for Case 1.

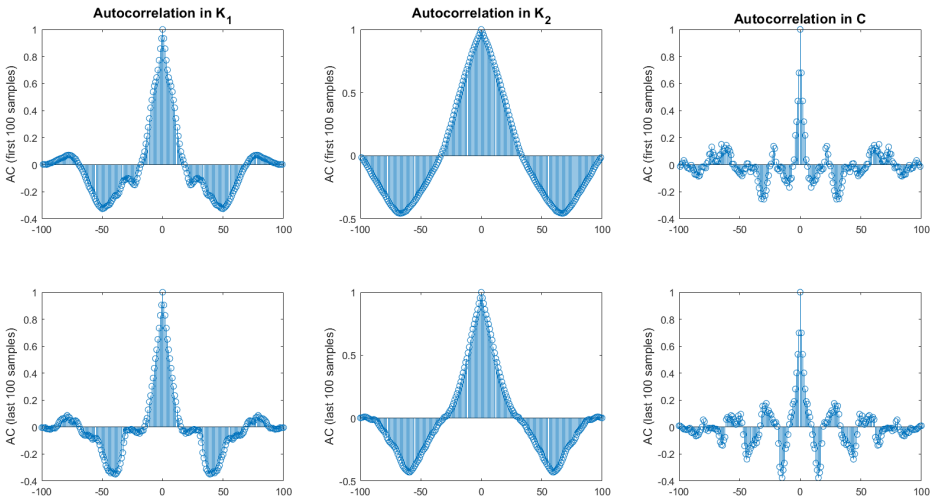


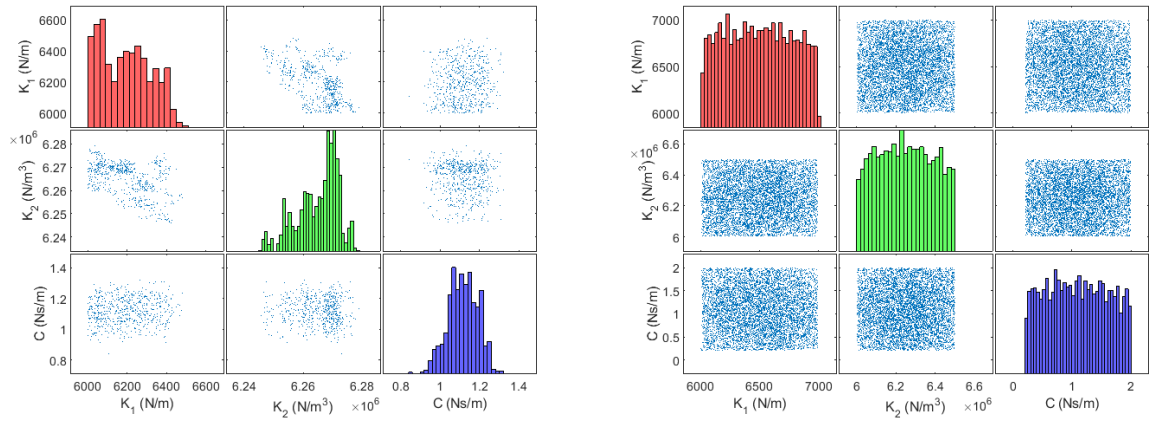
Figure 7: Autocorelation plot for K_1 and K_2 for first 100 and last 100 samples for Case 1.

Various approaches were considered for achieving stationary distribution and reducing the correlation among the samples.

- Burn-in period was increased to allow the chain to reach typical state and avoid the effect of initial value on sampling.
- Sample size i.e number of MCMC samples to be generated was increased (to 20000).
- Jump size in proposal distribution was varied by controlling the standard deviation of the proposal distribution
- Scaling was performed as the parameters were orders of magnitude different (varying from order 1e+3 to 1e+1) and the samplers perform best when all parameters are roughly on the same scale. Here, the scaling is implemented such that values remain in the range of +1 to -1.

Figure 8 shows the distribution of parameters for Case 1 before and after applying all the changes mentioned above. After scaling of parameters, the samples converged to mean values faster however, an increase in the

flat areas in the trace plot was seen resulting in a large number of rejected samples. From Figure 8, it can be seen that the distribution for K_2 and C are more identical to their true distribution.



(a) 5000 MCMC samples

(b) 20000 MCMC samples

Figure 8: Distribution of MCMC samples of parameters before (a) and after (b) implementing changes.

5.1 Results from MCMC sampling

5.1.1 Case 1: Cubic non-linearity

Parameters considered are K_1, K_2 and C. The distribution for K_1, K_2 and C (Figure. 9) are near to its true distribution which can be further ensured from KL divergence score in table 3 . Also, the range defined in true distribution for K_1, K_2 and C are covered in approximate distribution. The acceptance rate for the samples generated was found to be 0.33 which is an acceptable range for 1-d problem with three parameters. The distribution of C from MCMC samples (i.e approximate distribution) has a relatively larger divergence from its corresponding true distribution as indicated by a value of 0.81 as KL divergence score. However, the distribution is uniform with a range similar to the one defined for true distribution. Table 2 gives the range of the parameters considered for prior along with the range defined in measurement data (true distribution). Table 3 gives the mean and standard deviation of the distribution of the parameters (20000 MCMC samples).

Table 2: Range of parameters for true and prior distribution for Case 1.

Parameters	True		Prior	
	Minimum	Maximum	Minimum	Maximum
K_1 (N/m)	6000	7000	5400	7700
K_2 (N/m^3)	6000000	6500000	5400000	7150000
C (Ns/m)	0.2	2	0.1	5

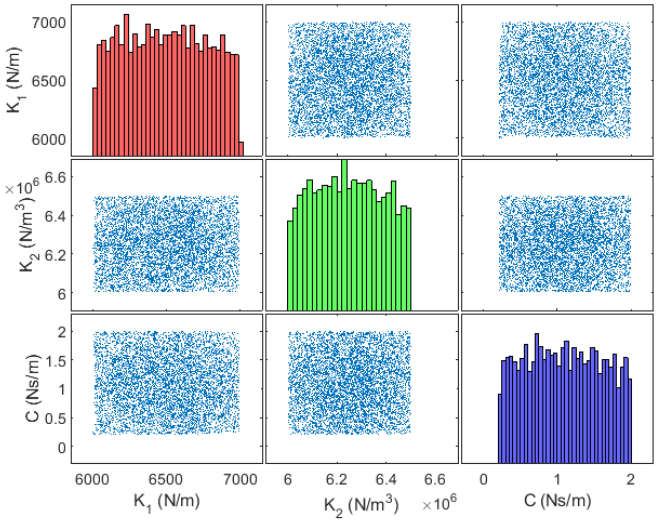


Figure 9: Scatter plot matrix of 20000 MCMC samples of K_1 (Red), K_2 (Green) and C (Navy blue) for Case 1.

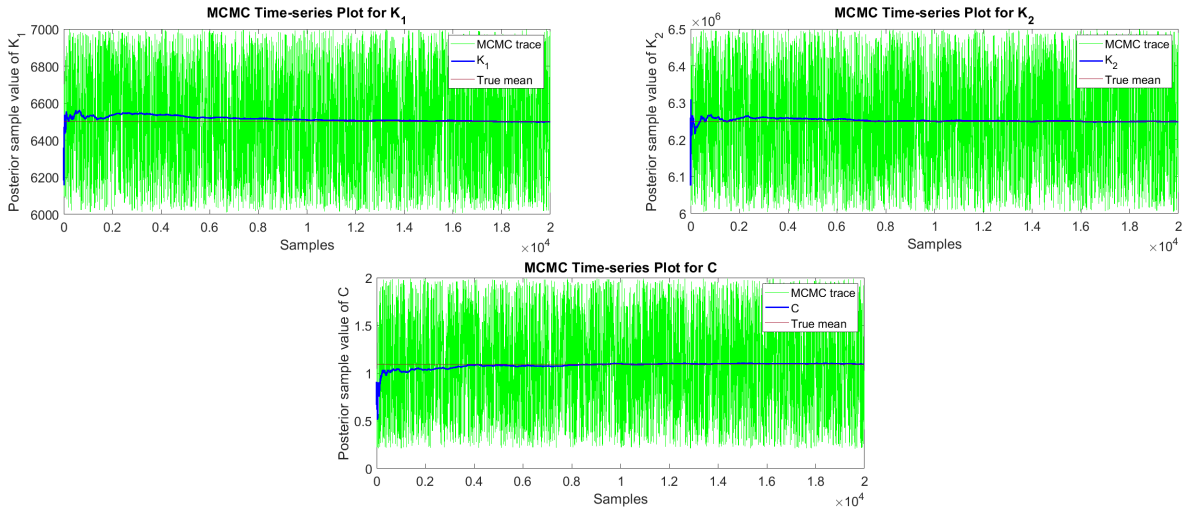


Figure 10: Trace plot for K_1, K_2 and C with 20000 MCMC samples

Table 3: Mean, standard deviation and KLD score of the distribution of parameters with 20000 samples for Case 1.

Parameters	Mean	Standard Deviation	KLD Score
K_1 (N/m)	6495.83	278.53	0.00
K_2(N/m^3)	6248829.85	138711.94	0.00
C (Ns/m)	1.09	0.52	0.81

5.1.2 Case 2: Quadratic damping with cubic stiffness

The distribution of the four parameters (K_1- Red, K_2- Green, C_1- Blue and C_2- Steel blue) of the system defined in Case 2 can be seen in Figure 11. K_1, K_2, C_1 and C_2 appear to have uniform distribution which matches the true distribution. From the trace plot (Figure 12), samples drawn from the posterior distribution

are almost perfect, the chain trace is approximately i.i.d (independent, identical distribution), and the acceptance rate is 0.23. The mean and standard deviation of the distribution of the parameters (20000 MCMC samples) are given in the table below (Table. 5).

Table 4: Range of parameters for true and prior distribution for Case 2.

Parameters	True		Prior	
	Minimum	Maximum	Minimum	Maximum
K_1 (N/m)	6000	7000	5400	7700
K_2 (N/m^3)	6000000	6500000	5400000	7150000
C_1 (Ns/m)	0.2	2	0.1	10
C_2 (Ns/m)	2	10	1	20

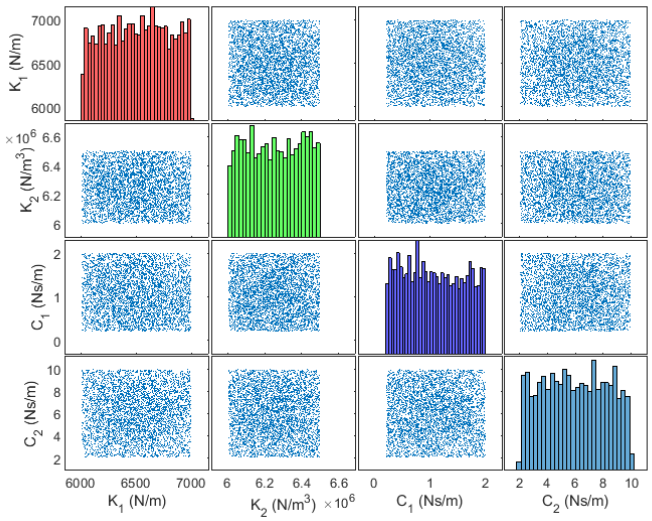


Figure 11: Scatter plot Matrix of 20000 MCMC samples of K_1 (Red), K_2 (Green), C_1 (Blue) and C_2 (Steel blue) for Case 2.

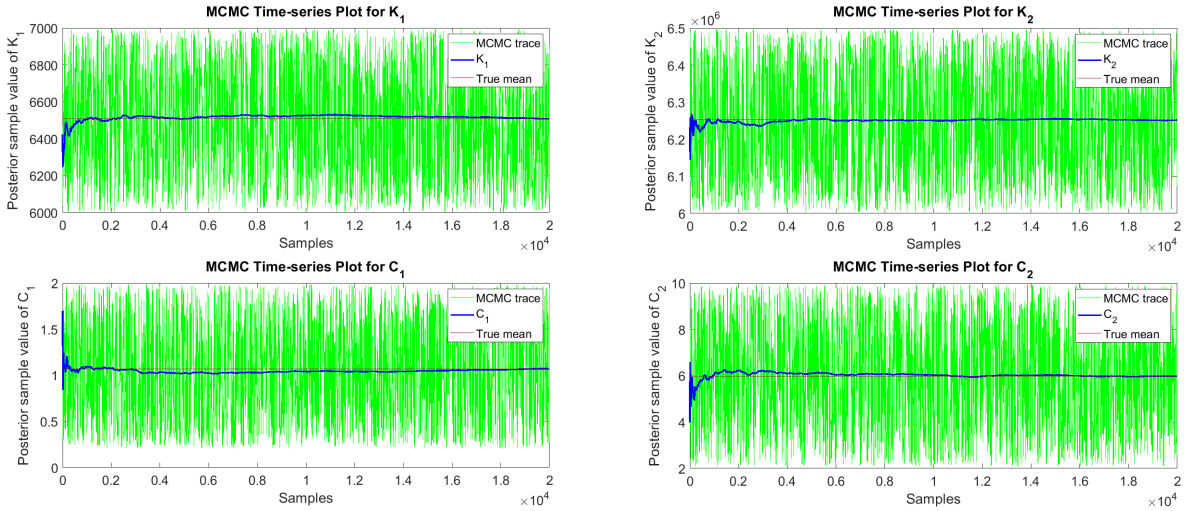


Figure 12: Trace plot for K_1, K_2, C_1 and C_2 with 20000 MCMC samples

Table 5: Mean, Standard deviation and KLD score of the distribution of parameters with 40000 samples for Case 2.

Parameters	Mean	Standard Deviation	KLD Score
K_1 (N/m)	6507.51	279.38	0.00
K_2 (N/m^3)	6253617.35	144994.10	0.00
C_1 (Ns/m)	1.08	0.52	0.81
C_2 (Ns/m)	5.97	2.26	0.00

5.1.3 Case 3: Dry friction

For Case 3, the range of parameters for true and prior distribution is defined in Table 6. Samples converge faster and uniform distribution of the parameters can be seen in Figure 13. From the scatterplot (Figure. 13), K, C_1 and C_2 samples appear to be uniformly distributed with a range similar to the one defined for true distribution. Low values KL divergence score for K, C_1 and C_2 show how near the approximated distribution is to the true distribution. The acceptance level of the MCMC sampler for Case 3 is 0.34. Table 7 gives the values of the mean, standard deviation and KL divergence score of the parameters.

Table 6: Range of parameters for true and prior distribution for Case 3.

Parameters	True		Prior	
	Minimum	Maximum	Minimum	Maximum
K (N/m)	6000	7000	5400	7700
C_1 (Ns/m)	0.1	0.9	0.1	5
C_2 (Ns/m)	0.2	2	0.1	10

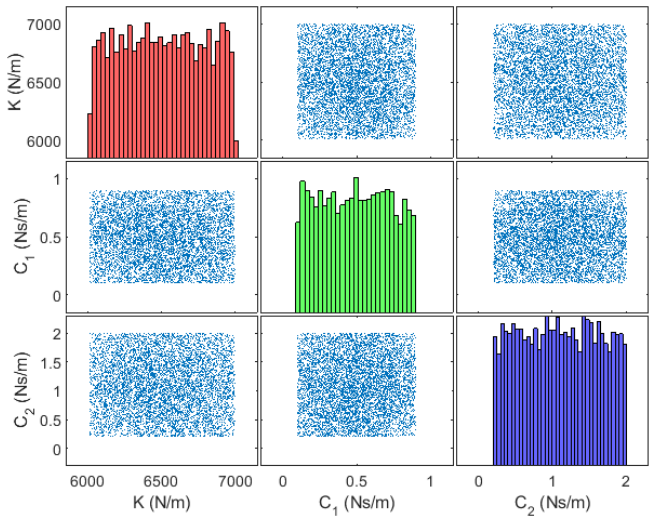


Figure 13: Scatter plot matrix of 20000 MCMC samples of K (Red), C_1 (Green)and C_2 (Blue) for Case 3.

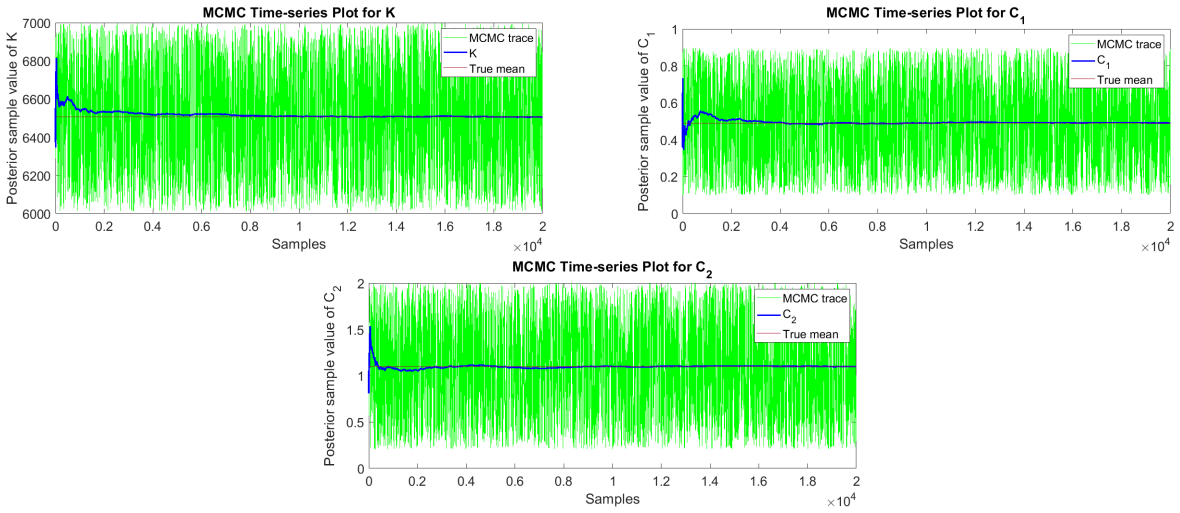


Figure 14: Trace plot for K, C_1 and C_2 with 20000 MCMC samples for Case 3.

Table 7: Mean, standard deviation and KLD score of the distribution of parameters with 20000 samples for Case 3.

Parameters	Mean	Standard Deviation	KLD Score
K (N/m)	6505.45	280.80	0.00
C_1 (Ns/m)	0.49	0.23	0.00
C_2 (Ns/m)	1.10	0.51	0.81

6 Conclusion

In this paper, stochastic nonlinear model updating is introduced based on using a measured backbone curve from identical structures that are assumed to be built in the same way but have variabilities due to manufacturing tolerances. The measured backbone curves are simulated using the proposed nonlinear models with stochastic parameters with known (true) initial parameters. The Likelihood function is defined by multiplying the fitted pdfs at different amplitudes of the backbone curve. Finally, the proposed method is applied to the three examples, i.e. cubic stiffness, quadratic damping with cubic stiffness and dry friction (Case1, 2 and 3) and the results showed the potential of the proposed method in identifying the distribution of updating parameters with a good degree of accuracy. Future work includes extending the analysis to other joint models comparing different methods of extracting backbone curves and their effect on parameter identification. Also, applying the proposed method using real experimental data of jointed structures.

Acknowledgements

This research work has been partly funded by the RCUK Energy Programme [grant number EP/T012250/1]. The views and options expressed herein do not necessarily reflect those of UKAEA. The support of the Supercomputing Wales project, which is part-funded by the European Regional Development Fund (ERDF) via the Welsh Government is also acknowledged.

References

[1] R. Ibrahim and C. Pettit, "Uncertainties and dynamic problems of bolted joints and other fasteners," *Journal of Sound and Vibration*, vol. 279, no. 3-5, pp. 857–936, Jan. 2005. [Online]. Available: https://doi.org/10.1016/j.jsv.2003.11.064

[2] T. L. Paez, L. J. Branstetter, and D. L. Gregory, "Modal randomness induced by boundary conditions," in *SAE Technical Paper Series*. SAE International, Oct. 1985. [Online]. Available: https://doi.org/10.4271/851930

[3] J. E. Mottershead and M. I. Friswell, "Model updating in structural dynamics: A survey," *Journal of Sound and Vibration*, vol. 167, no. 2, pp. 347–375, Oct. 1993. [Online]. Available: https://doi.org/10.1006/jsvi.1993.1340

[4] M. K. Vakilzadeh, "Stochastic model updating and model selection with application to structural dynamics," Ph.D. dissertation, Chalmers Publication Library, 2016.

[5] J. E. Mottershead, C. Mares, S. James, and M. I. Friswell, "Stochastic model updating: Part 2—application to a set of physical structures," *Mechanical Systems and Signal Processing*, vol. 20, no. 8, pp. 2171–2185, Nov. 2006. [Online]. Available: https://doi.org/10.1016/j.ymssp.2005.06.007

[6] K. P. Murphy, *Machine Learning: A Probabilistic Perspective*. London, England: MIT Press, 2012.

[7] C. Mares, B. Dratz, J. Mottershead, and M. Friswell, "Model updating using Bayesian estimation," vol. 5, 01 2006.

[8] H. Jensen and C. Vergara, "The use of Bayesian model updating in stochastic design problems," in *Proceedings of the 4th International Conference on Computational Methods in Structural Dynamics and Earthquake Engineering (COMPDYN 2013)*. Institute of Structural Analysis and Antiseismic Research School of Civil Engineering National Technical University of Athens (NTUA) Greece, 2014. [Online]. Available: https://doi.org/10.7712/120113.4507.c1421

[9] H. Sun and O. Büyüköztürk, "Bayesian model updating using incomplete modal data without mode matching," in *SPIE Proceedings*, T. Kundu, Ed. SPIE, Apr. 2016. [Online]. Available: https://doi.org/10.1117/12.2219300

[10] D. Wagg and S. Neild, *Nonlinear Vibration with Control*. Springer International Publishing, 2015. [Online]. Available: https://doi.org/10.1007/978-3-319-10644-1

[11] V. Ondra, I. Sever, and C. Schwingshackl, "A method for non-parametric identification of non-linear vibration systems with asymmetric restoring forces from a resonant decay response," *Mechanical Systems and Signal Processing*, vol. 114, pp. 239–258, Jan. 2019. [Online]. Available: https://doi.org/10.1016/j.ymssp.2018.05.010

[12] J. M. Londoño, S. A. Neild, and J. E. Cooper, "Identification of backbone curves of nonlinear systems from resonance decay responses," *Journal of Sound and Vibration*, vol. 348, pp. 224–238, Jul. 2015. [Online]. Available: https://doi.org/10.1016/j.jsv.2015.03.015

[13] A. Lye, A. Cicirello, and E. Patelli, "Sampling methods for solving Bayesian model updating problems: A tutorial," *Mechanical Systems and Signal Processing*, vol. 159, p. 107760, Oct. 2021. [Online]. Available: https://doi.org/10.1016/j.ymssp.2021.107760

[14] W. A. Link and M. J. Eaton, "On thinning of chains in MCMC," *Methods in Ecology and Evolution*, vol. 3, no. 1, pp. 112–115, Jun. 2011. [Online]. Available: https://doi.org/10.1111/j.2041-210x.2011.00131.x

[15] S. N. Maceachern and L. M. Berliner, "Subsampling the Gibbs sampler," *The American Statistician*, vol. 48, no. 3, pp. 188–190, 1994.