An efficient and robust sampler for Bayesian inference: Transitional Ensemble Markov Chain Monte Carlo

Adolphus Lye[a], Alice Cicirello[a,b], Edoardo Patelli[a,c,*]

[a] *Institute for Risk and Uncertainty, University of Liverpool*
[b] *Faculty of Civil Engineering and Geoscience, Delft University of Technology*
[c] *Centre for Intelligent Infrastructure, Department of Civil and Environmental Engineering, University of Strathclyde*

Abstract

Bayesian inference is a popular approach towards parameter identification in engineering problems. Such technique would involve iterative sampling methods which are often robust. However, these sampling methods often require significant computational resources and also the tuning of a large number of parameters. This motivates the development of a sampler called the Transitional Ensemble Markov Chain Monte Carlo.

The proposed approach implements the Affine-invariant Ensemble sampler in place of the classical Metropolis-Hastings sampler as the Markov chain Monte Carlo move kernel. In doing so, it allows for the sampling of badly-scaled and highly-anisotropic distributions without requiring extra computational costs. This makes the proposed sampler computationally efficient as a result of having less auxiliary parameters to compute per iteration compared to the standard single particle Transitional Markov Chain Monte Carlo. In addition to such change, an adaptive tuning algorithm is also proposed within the new sampler. This algorithm allows for automatic tuning of the step-size of the Affine-invariant Ensemble sampler. Hence, such proposals not only ensure that the new sampler is "tune-free" for the users, but also improves its robustness by ensuring that the acceptance rate of samples is well-controlled within accept-

*Corresponding author
Email address: edoardo.patelli@strath.ac.uk (Edoardo Patelli)

able bounds. As a result, this approach could be significantly faster compared to standard Transitional Markov Chain Monte Carlo methods on badly scaled and highly skewed distributions, which can be encountered when dealing with complex engineering problems.

The proposed sampler will be implemented on 2 benchmark numerical examples of varying complexities to demonstrate its strengths and advantages. In addition, the sampler is validated by investigating its parameter identification capability on an Aluminium Frame using experimental data.

Keywords: Bayesian inference, Model updating, Affine-invariant, Ensemble sampler, Transitional Markov Chain Monte Carlo, Structural health monitoring

---

1. Introduction

In recent years, Bayesian model updating has been increasingly employed to address numerous inverse problems in engineering due to its ability to combine prior knowledge with the observed data to yield a probabilistic description of the inferred parameter [1, 2]. For this reason, such approach has been adopted for parameter identification in numerous engineering structures including: 1) footbridges [3, 4, 5, 6], 2) buildings [7, 8, 9, 10, 11], and 3) aerospace designs [12, 13, 14]. Often, the Bayesian approach does not yield an analytical solution to the estimates of the inferred parameter(s) [15]. Hence, numerical techniques have been adopted in the form of sampling methods [16] such as Markov Chain Monte Carlo (MCMC) methods [15]. One such MCMC method is the Transitional Markov Chain Monte Carlo (TMCMC) sampler proposed by [17]. Recently, it has been applied in numerous engineering problems including: 1) characterizing the statistical uncertainties of the spatial variability parameters based on the Cone Penetration Test [18]; 2) analysing the multi-modality feature of the Bouc–Wen–Baber–Noori model of hysteresis [19]; 3) analysis of the creep behavior of soft soil and its associated uncertainty [20]; 4) executing reliability-based optimisation in linear structure designs subjected to random

excitations [21]; and 5) in estimating epistemic parameters within the DLR-AIRMOD structure [22]. Such vast application of the TMCMC highlights its popularity in engineering [23, 24]. This is attributed to its capability in estimating large number of parameters at one time (i.e. as many as 18 [15, 22]), sample from complex-shaped distributions [17, 19], and its ability to quantify the suitability of a model in describing the observed data under uncertainty [25, 26, 27, 28, 29].

Despite its strengths, the TMCMC algorithm presents numerous short-comings including: 1) having large number of auxiliary parameters to tune due to the choice of proposal distribution [30]; 2) computationally expensive due to the need to re-compute the auxiliary parameters at every iteration [15, 30]; 3) does not provide a mechanism to control the acceptance rates of the samples within the acceptable bounds of $[0.15, 0.50]$ [31]. To address those short-comings, an improved TMCMC approach, called the Transitional Ensemble Markov Chain Monte Carlo (TEMCMC) sampler, is presented. This approach employs the Affine-invariant Ensemble sampler (AIES) proposed by Goodman and Weare (2009) [32] in place of the Metropolis-Hastings (MH) sampler for the MCMC step. By using a sampler that satisfies the affine-invariance property, the performance of the method becomes independent from the complexity of the distribution [32]. To the best of the authors' knowledge, the proposed sampler has not been investigated yet in the literature. The motivations behind the proposed approach are: 1) the AIES can sample efficiently from highly-skewed distribution functions [32]; 2) it can be parallelised [33]; and 3) it removes the need of a proposal distribution, thereby reducing the number of auxiliary parameters to tune and the computational cost [32]. Moreover, an adaptive tuning algorithm is also proposed to automatically tune the step-size parameter of the AIES for every iteration. Thus, the objectives of the proposed TEMCMC sampler are to provide: 1) a sampler which is "tune-free" to the users; 2) a computationally less-expensive sampler than the existing TMCMC sampler, and 3) a robust method to effectively moderate acceptance rates within the acceptable bounds across the iterations.

The performance and the results obtained from the proposed algorithm will be compared and verified against the TMCMC. This will be done through the following numerical examples: 1) a Coupled Oscillator system where the objective is to infer 4 epistemic parameters (see Section 4.1); and 2) a 4-peaked Himmelblau's function where the objective is to observe the performance of each sampler in sampling from a multi-peaked posterior (see Section 4.2). Such comparisons will be done on the basis of the estimation of the inferred parameters, computational time elapsed in generating the posterior samples, and acceptance rates across the iterations j by the respective samplers. To validate the performance of the proposed algorithm under realistic settings, an application example in the form of the Aluminium Frame problem [34] is presented where the objective is to perform Bayesian model updating and infer the mass positions using actual experimental data (see Section 5).

2. Bayesian Model Updating

Bayesian model updating is based on the well-known concept of Bayesian inference [1, 2, 35] where prior assumptions or belief and the information available and/or measurements, are represented in terms of probability density functions. The posterior distribution on a set of parameters of interest is obtained according to [36]:

$$P(\boldsymbol{\theta}|\boldsymbol{D}, M) = \frac{P(\boldsymbol{D}|\boldsymbol{\theta}, M) \cdot P(\boldsymbol{\theta}|M)}{P(\boldsymbol{D}|M)} \tag{1}$$

where $\boldsymbol{\theta}$ represents the vector of uncertain parameters that is to be estimated, $\boldsymbol{D}$ represents the vector of the measurements (or observations) used in updating our knowledge of $\boldsymbol{\theta}$, and M represents the model (usually a function of $\boldsymbol{\theta}$) which is believed to best represent the available observations $\boldsymbol{D}$. The terms expressed in Eq. (1) are as such:

- $P(\boldsymbol{\theta}|M)$ is the prior distribution which describes our prior knowledge of $\boldsymbol{\theta}$ before any observation(s) is/are made,

- $P(\boldsymbol{D}|\boldsymbol{\theta}, M)$ is the likelihood function which reflects the degree of agreement and the error between $\boldsymbol{D}$ and the model output from M,
- $P(\boldsymbol{\theta}|\boldsymbol{D}, M)$ is the posterior distribution which describes our updated knowledge of $\boldsymbol{\theta}$ after observing the data $\boldsymbol{D}$,
- $P(\boldsymbol{D}|M)$ is the evidence which serves as the normalising constant of the posterior.

Due to $P(\boldsymbol{D}|M)$ being a numerical constant which is independent of $\boldsymbol{\theta}$, it can be neglected and the posterior is re-expressed up to a normalising constant:

$$P(\boldsymbol{\theta}|\boldsymbol{D}, M) \propto P(\boldsymbol{D}|\boldsymbol{\theta}, M) \cdot P(\boldsymbol{\theta}|M) \tag{2}$$

Details to the above terms can be found in [15].

Due to the nature of $P(\boldsymbol{\theta}|\boldsymbol{D}, M)$ in Eq. (2), standard Monte Carlo approaches become inapplicable as they require that the analytical function of the distribution to be normalised and that they have a defined Cumulative Distribution Function (CDF) [37, 38]. For these reasons, advanced sampling techniques have been developed to sample from unnormalised distributions including: 1) Markov Chain Monte Carlo (MCMC) [39], 2) Sequential Monte Carlo (SMC), and 3) Transitional Markov Chain Monte Carlo (TMCMC) samplers [40]. Detailed explanations to each of these techniques are provided in [15].

2.1. Review of the Transitional Markov Chain Monte Carlo

TMCMC is one of the most successfully applied sampling techniques in engineering applications [23, 24]. It is based on the Adaptive Metropolis-Hastings (AMH) technique [41] whereby samples are obtained from a series of intermediate "transitional" distributions P^j defined as [17]:

$$P^j \propto P(\boldsymbol{D}|\boldsymbol{\theta}, M)^{\beta_j} \cdot P(\boldsymbol{\theta}|M) \tag{3}$$

Here, j denotes the transition step number taking values between 0 to m, where m indicates the last iteration number, β_j is the tempering parameter which takes values such that $\beta_0 = 0 < \beta_1 < \cdots < \beta_{m-1} < \beta_m = 1$. In

doing so, this allows for the transitional distribution to transit from the prior to the posterior distribution (i.e. $P^0 = P(\boldsymbol{\theta}|M)$ to $P^m = P(\boldsymbol{\theta}|\boldsymbol{D}, M)$) in an iterative manner. A critical aspect of TMCMC is the choice of the step-size $\Delta\beta_j = \beta_j - \beta_{j-1}$ to ensure a smooth and gradual transition from $P(\boldsymbol{\theta}|M)$ to $P(\boldsymbol{\theta}|\boldsymbol{D}, M)$ [17]. In addition, the magnitude of $\Delta\beta_j$ itself has a direct impact on the acceptance-rates of the samples [15]. A large step-size leads to large number of rejected proposed samples due to an increased likelihood of these samples lying outside the support of the posterior (whilst still within the support of the prior). A small $\Delta\beta_j$ leads to an increase in the number of iterations required. The optimal value of $\Delta\beta_j$ is such that the Coefficient Of Variation (COV) of the nominal weights of the samples $\boldsymbol{\theta}_i$, defined as [17]:

$$w_i^j = P(\boldsymbol{D}|\boldsymbol{\theta}_i, M)^{\Delta\beta_j} \tag{4}$$

is close to 1 for any given iteration j. As such, the optimal tempering factor β_j for each iteration can then be determined as [17, 30]:

$$\arg\max_{\beta_j}\{|\sigma(w^j)/\mu(w^j) - 1|\} \tag{5}$$

.

At iteration $j = 0$, the TMCMC algorithm is initialized by generating N samples from the prior $P(\boldsymbol{\theta}|M)$ via direct Monte Carlo sampling. Next, at iteration $j = 1$ (i.e. while $\beta_j < 1$), the nominal weight of the samples is computed according Eq. (4) and normalised by dividing the weights by its summation (i.e. $\sum_{i=1}^{N} w_i^j$). Let the normalised weights be denoted as $\hat{w}_i^j$. Following which, N single-step Markov Chains are initiated. To determine the starting samples for each of the Markov Chains, N samples of $\boldsymbol{\theta}_i$ are resampled (with replacement) via weighted resampling. This is done such that $\boldsymbol{\theta}_i$ is sampled with probability $\hat{w}_i^j$ [42]. The MH approach is then used to generate 1 sample from each of those Markov Chains [39]: candidate samples $\boldsymbol{\theta}^*$ are sampled from a proposal distribution $q(\boldsymbol{\theta}^*|\boldsymbol{\theta})$ with mean $\bar{\boldsymbol{\theta}}^j$ and covariance matrix $\boldsymbol{\Sigma}^j$ defined as:

$$\bar{\boldsymbol{\theta}}^j = \sum_{i=1}^{N} \boldsymbol{\theta}_i \cdot \hat{w}_i^j \tag{6}$$

and

$$\boldsymbol{\Sigma}^j = \gamma^2 \cdot \sum_{i=1}^{N} \hat{w}_i^j \cdot \left[\{\boldsymbol{\theta}_i - \bar{\boldsymbol{\theta}}\} \times \{\boldsymbol{\theta}_i - \bar{\boldsymbol{\theta}}\}^T \right]. \tag{7}$$

γ is the scaling parameter of $\boldsymbol{\Sigma}^j$. The i^{th} candidate sample is accepted with probability α_i:

$$\alpha_i = min\left[1, \frac{P^j(\boldsymbol{\theta}_i^*)}{P^j(\boldsymbol{\theta}_i)}\right] \tag{8}$$

where $P^j(\boldsymbol{\theta}_i^*)$ is the density value of P^j computed at $\boldsymbol{\theta}_i^*$ while $P^j(\boldsymbol{\theta}_i)$ is the density value of P^j computed at $\boldsymbol{\theta}_i$. The values of the new tempering parameter is updated and the algorithm repeats the described procedure for subsequent iterations $j = j + 1$ till $\beta_j = 1$.

The TMCMC sampler computes the evidence term $P(\boldsymbol{D}|M)$ of the posterior distribution [17, 30]. The metric $P(\boldsymbol{D}|M)$ quantifies how well a given model M describes the available observations $\boldsymbol{D}$. The $P(\boldsymbol{D}|M)$ is estimated as the product of the mean of the nominal weights w_i^j at any given iteration j:

$$P(\boldsymbol{D}|M) \approx \prod_{j=1}^{m} \frac{1}{N} \cdot \sum_{i=1}^{N} w_i^j \tag{9}$$

2.2. Limitations of the approach

One key problem in the TMCMC technique is that there is no fixed universal value for the scaling parameter γ although it was stated in [17] that the " optimal" value would be 0.2. This was highlighted in [30] where it was argued that this " optimal" value is not applicable for all cases. In fact, different values for γ have been utilised in different research such as 0.5 in [21] and 1.0 in [18]. Hence, [30] proposed a tuning algorithm to adaptively adjust γ. Starting from an initial value of $\gamma^{j=1} = \frac{2.4}{\sqrt{N_d}}$ [38, 43] where N_d is the dimension of $\boldsymbol{\theta}$. Upon the conclusion of the MCMC step in updating the samples, the mean acceptance rate for the current iteration α^j is obtained. This mean acceptance rate is then compared against the target acceptance rate α_{tr} which is defined as [31]:

$$\alpha_{tr} = \frac{0.21}{N_d} + 0.23 \tag{10}$$

Once this is done, the scale parameter is then tuned and updated according to:

$$\gamma^{j+1} = \gamma^j \cdot exp\left[\frac{\alpha^j - \alpha_{tr}}{j}\right] \tag{11}$$

Such approach should also help to moderate the overall acceptance rate of the samples such that α^j falls between 0.2 and 0.5 as much as possible [31]. However, to the best of the authors' knowledge, investigations into the acceptance rates of the TMCMC sampler across iterations have not been done. The TMCMC sampling procedure is summarized in Algorithm 1.

Algorithm 1 TMCMC sampler algorithm

procedure (Generate N samples from $P(\boldsymbol{\theta}|\boldsymbol{D}, M)$)
  Set $j = 0$ and $\beta_j = 0$ ▷ Initialise iteration counters
  Set $\gamma^{j+1} = \frac{2.4}{\sqrt{N_d}}$ ▷ Initialise scale parameter
  Draw initial N sample set: $\boldsymbol{\theta}_i \sim P(\boldsymbol{\theta}|M)$ ▷ Generate samples from the prior
  while $\beta_j < 1$ **do** ▷ Main sampling loop
    Set $j = j + 1$
    Compute $\Delta\beta_j$ using Eq. (5)
    Compute P^j using Eq. (3)
    for $i = 1 : N$ **do** ▷ For each i^{th} chain (MCMC step)
      Resample: $\boldsymbol{\theta}_i \sim \hat{w}_i^j$
      Draw candidate sample: $\boldsymbol{\theta}^* \sim q(\boldsymbol{\theta}_i^*|\boldsymbol{\theta}_i)$
      Accept/Reject $\boldsymbol{\theta}^*$ with probability α_i using Eq. (8)
    end for
    Compute γ^{j+1} using Eq. (11)
  end while
  Compute $P(\boldsymbol{D}|M)$ using Eq. (9)
end procedure

Current TMCMC sampler is also computational inefficient since the mean $\bar{\boldsymbol{\theta}}^j$ and covariance matrix $\boldsymbol{\Sigma}^j$ need to be calculated at each iteration j. This adds

additional parameters that need to be adaptively tuned, on top of β_j and γ^j, as a result of the choice of proposal distribution $q(\boldsymbol{\theta}^*|\boldsymbol{\theta}_i)$ used. Furthermore, there is a loose assumption that it is optimal to set $q(\boldsymbol{\theta}^*|\boldsymbol{\theta}_i)$ as a Normal distribution [39]. Strictly speaking, such choice only provides convenience in simplifying the sampling process of $\boldsymbol{\theta}^*$ from $P(\boldsymbol{\theta}|\boldsymbol{D}, M)$ and the computation of the acceptance ratio α. Such convenience is attributed to the symmetrical nature of the Normal distribution [39, 44]. In fact, there are specific cases whereby asymmetric proposal distributions are required. For instance, if one is to estimate the posterior of a variance parameter, to ensure that the proposed candidate samples $\boldsymbol{\theta}^*$ are never less than 0, the proposal distribution should be such that it is skewed towards positive values (i.e. $\boldsymbol{\theta}^* > 0$) such as the Log-normal distribution [45]. This would ensure that the overall acceptance rates α would not be too low (i.e. $\alpha < 0.15$) [46]. Optimally, the proposal distribution should follow that of the posterior [44] but such approach is not feasible due to the lack of apriori knowledge over the analytical form of the actual posterior itself. Thus, the choice of $q(\boldsymbol{\theta}^*|\boldsymbol{\theta}_i)$ adds some degree of uncertainty.

3. Transitional Ensemble Markov Chain Monte Carlo

Numerous alternative MCMC kernels in TMCMC have been considered in previous studies including: 1) Slice sampler (i.e. TMCMC-Slice) [47, 48], 2) Metropolis-Adjusted-Langevin (MAL) sampler (i.e. L-TMCMC) [49], and 3) Differential evolution MCMC kernel (i.e. DE-TMCMC) [50]. A summary to each of the above TMCMC variants is provided.

The TMCMC-Slice is able to produce estimates with significantly lower COV than that of TMCMC. This is because the Slice sampler algorithm [51] draws samples more efficiently than MH from P^j such that candidate samples $\boldsymbol{\theta}_i^*$ are now drawn closer to regions of higher probability in $P(\boldsymbol{\theta}|\boldsymbol{D}, M)$. This leads to a higher convergence of $\boldsymbol{\theta}_i$ and a smaller spread about the sample space defined by $P(\boldsymbol{\theta}|\boldsymbol{D}, M)$ [47]. In addition, the use of the Slice sampler algorithm removes the need for $q(\boldsymbol{\theta}^*|\boldsymbol{\theta}_i)$ thereby removing the need to compute $\bar{\boldsymbol{\theta}}^j$ and $\boldsymbol{\Sigma}^j$ [51, 48].

A significant drawback of the TMCMC-Slice is its relatively long computational time compared to the TMCMC due to the computation cost involved in tuning the auxiliary variables and step-size of the Slice sampler [48].

The L-TMCMC demonstrated high parallel efficiency by adopting an adaptive kriging metamodel, in place of the true model, to perform the model evaluations when computing the likelihood function [49]. This reduces computation time, thereby making the algorithm computationally less-expensive compared to the TMCMC [49]. In addition, due to the use of the MAL move kernel, it allows for the L-TMCMC to sample more efficiently from multi-modal $P(\boldsymbol{\theta}|\boldsymbol{D}, M)$ as well as a better identification of parameters, especially those which were unidentifiable by TMCMC [49]. However, its performance suffers when $P(\boldsymbol{\theta}|\boldsymbol{D}, M)$ is highly-dimensional due to the computation of the gradient information by the MAL algorithm which is computationally ineffective and inefficient under such settings [49].

The DE-TMCMC demonstrated high parallel efficiency due to the use of the Differential evolution kernel, thereby making it computationally more efficient compared to the TMCMC [50]. In addition, the sampler is able to produce estimates with higher degree of accuracy due to its maximum a posteri estimates being closer to the true values of the inferred parameter(s) compared to the TMCMC [50]. However, the sampler still has significant number of auxiliary parameters to tune for the differential evolution MCMC kernel. This could add significant computational costs [50].

In this work, we implement the AIES as the alternative MCMC kernel to the MH. The reasons for this are the following: 1) to exploit the efficiency of the AIES in its ability to sample from highly-skewed and anisotropic distributions; 2) the AIES can be parallelised; and 3) no proposal distribution is required which, in turn, reduces the number of auxiliary parameters to tune. This gives rise to the TEMCMC to which an additional feature proposed is the adaptive tuning algorithm which automatically tunes its step-size parameter and moderate the sampler's acceptance rates.

3.1. Review of the Affine-invariant Ensemble Sampler

The AIES is a MCMC sampling technique recently developed which possesses the affine-invariance property [32]. Currently, it has been applied across numerous research fields such as Cosmology [33, 52, 53, 54, 55, 20], Physics [56, 57, 58], and Engineering [59, 60, 61, 62, 63]. To provide an understanding of the AIES sampler, this section will first present the concept of affine-transformation, followed by a definition and description of an ensemble before finally, explaining the sampler.

An affine-transformation operation ψ is an invertible linear mapping from a $\mathbb{R}^{N_d}$ to $\mathbb{R}^{N_d}$ space [64]:

$$\psi(\boldsymbol{\theta}) : \boldsymbol{\Theta} = \hat{A}\, \boldsymbol{\theta} + \boldsymbol{b} \tag{12}$$

where $\boldsymbol{\Theta}$ represents $\boldsymbol{\theta}$ in the affine-transformed space, $\hat{A}$ is the N_d-by-N_d non-singular transformation matrix and $\boldsymbol{b}$ is the N_d-by-1 translation vector. This simple transformation transforms a difficult sampling problem into a tractable one, in a similar way in which a deterministic optimization problem is transformed to deal with a well-scaled function [32]. Let $P'(\boldsymbol{\Theta}|\boldsymbol{D}, M)$ represents a general class of densities describing the posterior distribution of $\boldsymbol{\Theta}$ where P' denotes the distribution function in the $\boldsymbol{\Theta}$-space. If the following condition holds [32]:

$$P'(\boldsymbol{\Theta}|\boldsymbol{D}, M) = P'(\psi(\boldsymbol{\theta})|\boldsymbol{D}, M) \propto P(\boldsymbol{\theta}|\boldsymbol{D}, M) \tag{13}$$

then, $\boldsymbol{\Theta}$-space and $\boldsymbol{\theta}$-space are said to be affine-invariant [65]. Therefore, the Affine-invariant sampler can be constructed by using a proposal distribution q' with the form [32]:

$$q'(\boldsymbol{\Theta}^*|\boldsymbol{\Theta}_i) = q'(\psi(\boldsymbol{\theta}^*)|\psi(\boldsymbol{\theta}_i)) \propto q(\boldsymbol{\theta}^*|\boldsymbol{\theta}_i) \tag{14}$$

This proposal distribution is also invariant under affine-transformation, i.e. the probability of generating a sample $\boldsymbol{\Theta}^*$ given $\boldsymbol{\Theta}_i$ in the transformed $\boldsymbol{\Theta}$-space

is equal to the probability of generating a sample $\boldsymbol{\theta}^*$ given $\boldsymbol{\theta}_i$ in the original $\boldsymbol{\theta}$-space [32].

Within the AIES, a collection of N_c Markov chains: $\vec{\boldsymbol{\theta}}_i = \{\boldsymbol{\theta}_{1,i}, \boldsymbol{\theta}_{2,i}, \ldots, \boldsymbol{\theta}_{N_c-1,i}, \boldsymbol{\theta}_{N_c,i}\}$ is first generated. The proposal sample $\boldsymbol{\theta}_k^*$ is then obtained by using a move kernel that is invariant towards an affine-transformation and that uses the current sample and that obtained from a complementary chain. In practice, it is required that the number of chains N_c be at least twice the dimension of $\boldsymbol{\theta}$ (i.e. $N_c \geq 2 \times N_d$) [32]. Each chain generates 1 sample from the prior giving the first ensemble $\vec{\boldsymbol{\theta}}_i$ for $i = 1$. Once this is done, the samples in the ensemble are then updated one chain at a time. To update the k^{th} chain (for $k = 1, \ldots, N_c$), a sample from a complementary chain is randomly selected from the set $\vec{\boldsymbol{\theta}}_{[k],i} = \{\boldsymbol{\theta}_{1,i+1}, \ldots, \boldsymbol{\theta}_{k-1,i+1}, \boldsymbol{\theta}_{k+1,i}, \ldots, \boldsymbol{\theta}_{N_c,i}\}$. Let this chosen sample from the set $\vec{\boldsymbol{\theta}}_{[k],i}$ be denoted as $\boldsymbol{\theta}_{[k]}$. Following which, the candidate sample for the k^{th} chain $\boldsymbol{\theta}_k^*$ is generated. This can be expressed as [32, 33]:

$$\boldsymbol{\theta}_k^* = \boldsymbol{\theta}_{[k]} + \lambda \cdot (\boldsymbol{\theta}_{k,i} - \boldsymbol{\theta}_{[k]}) \tag{15}$$

whereby λ is real-valued scalar proposal stretch factor of the stretch-move affine-transformation [32, 33]. λ can be represented as a random variable following a proposal distribution $g(\lambda)$ [32]. Analogous to the MH algorithm where a symmetric $q(\boldsymbol{\theta}^*|\boldsymbol{\theta}_i)$ is used, $g(\lambda)$ is chosen to satisfy the symmetry condition such that [32]:

$$g\left(\frac{1}{\lambda}\right) = \lambda \cdot g(\lambda) \tag{16}$$

so that the stretch-move in Eq. (15) is symmetric [32]. For this reason, $g(\lambda)$ is proposed in [32] as:

$$g(\lambda) = \begin{cases} \frac{1}{2\cdot(\sqrt{u}-\frac{1}{\sqrt{u}})} \cdot \frac{1}{\sqrt{\lambda}} & \text{if } \lambda \in [\frac{1}{u}, u] \\ 0 & \text{otherwise} \end{cases} \tag{17}$$

whereby u serves as the user-defined step-size of the AIES sampler which needs to be strictly greater than 1. Once $\boldsymbol{\theta}_k^*$ is sampled following Eq. (15), it is then

accepted with probability α_k:

$$\alpha_k = min\left[1,\ \lambda^{N_d-1} \cdot \frac{P(\boldsymbol{\theta}_k^*|\boldsymbol{D}, M)}{P(\boldsymbol{\theta}_{k,i}|\boldsymbol{D}, M)}\right] \tag{18}$$

Once the samples in all N_c chains have been updated, set $i = i + 1$ and the updating procedure repeats itself until $i = N$. In summary, the entire sampling procedure by AIES is summarized and illustrated in Algorithm 2. Interested readers will find further theoretical and numerical investigations of ensemble samplers with affine invariance properties in reference [32].

Figure 1 illustrates the stretch-move in a $2D$ sample space $\boldsymbol{\theta} = \{\theta^1, \theta^2\}$. In the figure, the sample of the 3^{rd} chain (in red) is being updated. The complementary sample for this case is $\boldsymbol{\theta}_{4,i}$. A straight line initiating from $\boldsymbol{\theta}_{4,i}$ is drawn to include $\boldsymbol{\theta}_{3,i}$. This straight line is the path along which the candidate sample $\boldsymbol{\theta}_3^*$ could possibly lie. By sampling λ from $g(\lambda)$, and using Eq. (15), $\boldsymbol{\theta}_3^*$ is defined and represented in blue.

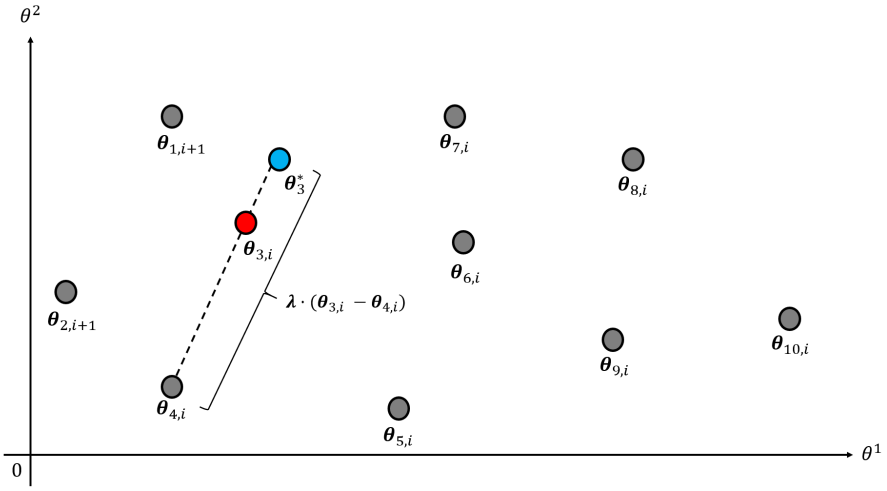


Figure 1: Schematic diagram of the stretch-move that is used to update the sample of the 3^{rd} chain in red. Here, the candidate sample for the 3^{rd} chain is represented in blue while the randomly chosen complementary sample in this case is that from the 4^{th} chain. Image adapted from [32].

(19)

Algorithm 2 AIES sampler algorithm

procedure (Generate N samples from $P(\boldsymbol{\theta}|\boldsymbol{D}, M)$)
  Define N_c chains: $\vec{\boldsymbol{\theta}}_1 = \{\boldsymbol{\theta}_{1,1}, \boldsymbol{\theta}_{2,1}, ..., \boldsymbol{\theta}_{N_c-1,1}, \boldsymbol{\theta}_{N_c,1}\}$
  for $i = 1 : N - 1$ **do**
    for $k = 1 : N_c$ **do**
      Select randomly $\boldsymbol{\theta}_{[k]}$ from set $\vec{\boldsymbol{\theta}}_{[k],i}$
      Sample: $\lambda \sim g(\lambda)$
      Generate $\boldsymbol{\theta}_k^*$ using Eq. (15)
      Calculate acceptance probability α_{AIES} using Eq. (19)
      Sample: $r \sim U[0, 1]$
      if $\alpha_k > r$ **then**
        Set $\boldsymbol{\theta}_{k,i+1} = \boldsymbol{\theta}_k^*$
      else
        Set $\boldsymbol{\theta}_{k,i+1} = \boldsymbol{\theta}_{k,i}$
      end if
    end for
  end for
end procedure

The affine-invariant property of the stretch-move is explained as such. Supposed 2 independent sampling procedures are conducted by the AIES: one to sample $\boldsymbol{\theta}$ from $P(\boldsymbol{\theta}|\boldsymbol{D}, M)$), and the other to sample $\boldsymbol{\Theta}$ from $P'(\boldsymbol{\Theta}|\boldsymbol{D}, M)$). Given the same sequence of λ_i in both runs for $i \geq 2$, and that the starting samples $\boldsymbol{\Theta}_1$ and $\boldsymbol{\theta}_1$ are related according to Eq. (19), the AIES is able to generate sample sequences such that the relationship between $\boldsymbol{\Theta}_i$ and $\boldsymbol{\theta}_i$ is always upheld for all iterations i. The mathematical illustration of the affine-invariant stretch-move is provided by Eq. (15). By performing an affine-transformation

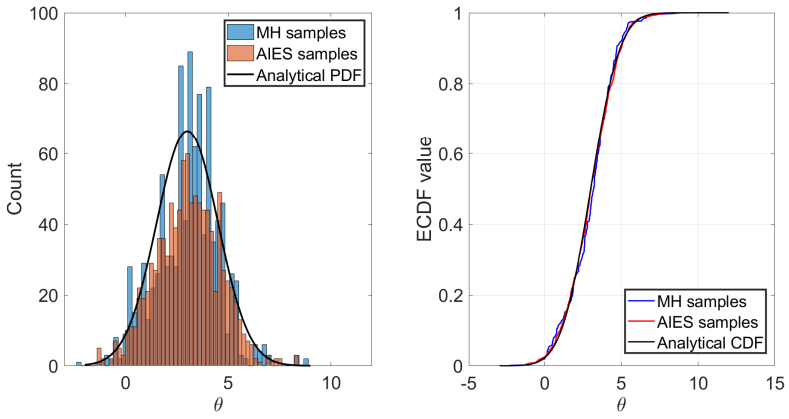


Figure 2:

on both sides of the equation, we obtain [32]:

$$\boldsymbol{\Theta}_k^* : \psi(\boldsymbol{\theta}_k^*) = \psi(\boldsymbol{\theta}_{[k]}) + \lambda_i \cdot (\psi(\boldsymbol{\theta}_{k,i}) - \psi(\boldsymbol{\theta}_{[k]})) \tag{20}$$

Expanding and re-arranging the above equation, one will obtain the final expression [32]:

$$\boldsymbol{\Theta}_{[k]} + \lambda_i \cdot (\boldsymbol{\Theta}_{k,i} - \boldsymbol{\Theta}_{[k]}) = \hat{A}\left[\boldsymbol{\theta}_{[k]} + \lambda_i \cdot (\boldsymbol{\theta}_{k,i} - \boldsymbol{\theta}_{[k]})\right] + \boldsymbol{b} \tag{21}$$

Eq. (21) implies 2 key things: 1) the generation of samples from the affine-transformed $\boldsymbol{\Theta}$-space is no different from sampling from the original $\boldsymbol{\theta}$-space up to an affine-transformation [32]; and 2) the probability of sampling $\boldsymbol{\Theta}_k^*$ starting from $\boldsymbol{\Theta}_{k,i}$ in the $\boldsymbol{\Theta}$-space is equal to the probability of sampling $\boldsymbol{\theta}_k^*$ starting from $\boldsymbol{\theta}_{k,i}$ in the original $\boldsymbol{\theta}$-space. The second point satisfies the condition defined by Eq. (14) which endows the stretch-move its affine-invariant property.

Hence, one key advantage of the AIES sampler over the MH sampler is its ability to sample from a poorly-scaled and highly-anisotropic distributions just as effectively and efficiently as it would from a well-scaled affine-transformed distribution [32, 66]. To illustrate this, a numerical example is presented in Section 7 to which readers can refer to.

Another advantage that the AIES sampler has over the MH sampler is the absence of any user-defined $q(\boldsymbol{\theta}^*|\boldsymbol{\theta}_i)$ to generate candidate samples. This reduces the degree-of-freedom as well as the number of parameters to be adaptively tuned by 2 given that the sample mean $\bar{\boldsymbol{\theta}}$ and its covariance matrix $\boldsymbol{\Sigma}^j$ need not be computed at every j. This leaves β_j and u as the only parameters to be adaptively tuned, thereby improving the computational efficiency for the TEMCMC sampler.

As such, this motivates the implementation of the AIES algorithm in the proposed TEMCMC sampler given that the "transitional" distributions P^j (see Eq. (3)) can be highly-skewed and anisotropic in general and its sampling performance would be least affected by the scaling of P^j across the transition step j.

3.2. Adaptive-tuning algorithm

To adaptively tune the step-size parameter u, an algorithm is proposed based on the work by [30]. The initial step-size value $u^{j=1}$ is set at 2 given that this is the "optimal" value for most problems [33, 52]. From this initial value, the nominal step-size u_{nom} is computed after the MCMC step:

$$u_{nom} = u^j \cdot exp\left[\alpha^j - \alpha_{tr}\right] \tag{22}$$

where α^j is the acceptance rate for the current iteration and α_{tr} is the target acceptance rate (see Eq. (10)). The acceptance rate α^j is treated as a random variable. Consequently, the nominal step-size u_{nom} is also a random variable which can be randomized and adapted through α^j. Such implementation is not limited to AIES, but is a generic property of MCMC samplers [30, 67]. If $u_{nom} > 1$, then $u^{j+1} = u_{nom}$. Otherwise, the algorithm sets $u^{j+1} = 1.01$ to ensure that the step-size would never fall below 1. This procedure is repeated at the end of every iteration until the last transition step $j = m$.

In summary, the proposed TEMCMC sampler possesses 3 key benefits: 1) it is practically "tune-free" for the users; 2) it is computationally less expensive compared to TMCMC sampling; and 3) its acceptance rate is moderated such

that it falls within the acceptance range between 0.15 and 0.50 [46] for the majority of the transition steps j. Such benefits and strengths of the TEMCMC sampler will be illustrated in the numerical examples as well as the Aluminium Frame application problem in Sections 4 and 5 respectively. A pseudoalgorithm of the TEMCMC sampler is provided in Algorithm 3.

Algorithm 3 Proposed TEMCMC sampler algorithm

procedure (Generate N samples from $P(\boldsymbol{\theta}|\boldsymbol{D}, M)$)
  Set $j = 0$ and $\beta_j = 0$ ▷ Initialise
  Draw N initial sample set: $\boldsymbol{\theta}_i \sim P(\boldsymbol{\theta}|M)$
  Set $a^{j+1} = 2$ ▷ Set initial value of step-size
  while $\beta_j < 1$ **do** ▷ Main sampling loop
    Set $j = j + 1$
    Compute $\Delta\beta_j$ using Eq. (5)
    Compute P^j using Eq. (3)
    Resample N samples: $\boldsymbol{\theta}_i \sim \hat{w}_i^j$
    Set $\boldsymbol{\theta}_i = \boldsymbol{\theta}_{i,1}$ in ensemble $\vec{\boldsymbol{\theta}}_1$ ▷ Initiate ensemble
    Update $\vec{\boldsymbol{\theta}}_1$ with 1 iteration of AIES (see Algorithm 2) ▷ MCMC step
    Set samples in ensemble $\vec{\boldsymbol{\theta}}_2$ as samples of P^j
    Compute u_{nom} using Eq. (22) ▷ Tuning the step-size
    if $a_{nom} > 1$ **then**
      Set $u^{j+1} = u_{nom}$
    else
      Set $u^{j+1} = 1.01$
    end if
  end while
  Compute $P(\boldsymbol{D}|M)$ using Eq. (9)
end procedure

4. Numerical Examples

4.1. 2-DOF Coupled Oscillator System

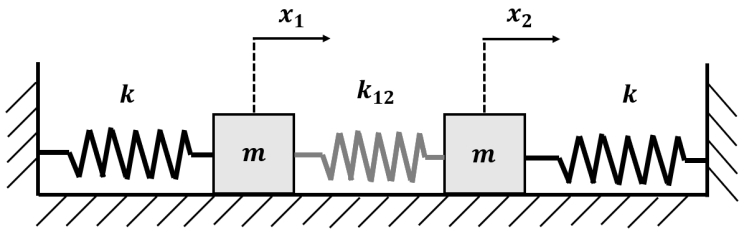


Figure 3: Schematic diagram of the 2-*DOF* Coupled Oscillator system based on the set-up in [68].

The objective of this numerical example is to observe and compare the differences in the performance of the TMCMC sampler and the proposed TEMCMC sampler in a 4-dimensional Bayesian model updating set-up. This comparison will be done on the basis of the COV of the estimation of the epistemic parameters $\boldsymbol{\theta}$, the computation time elapsed in sampling from the posterior, and the acceptance rates across the transition steps j.

Figure 3 illustrates a simple 2 Degrees-of-Freedom (DOF) coupled oscillator set-up consisting of 2 equal-sized blocks with equal mass m attached to primary springs with stiffness k and an inter-mass secondary spring with stiffness k_{12}. x_1 and x_2 denote the respective instantaneous displacement of the blocks. In this problem, the mass of the blocks are fixed at $m = 0.5\ kg$. In addition, it is assumed that both k and k_{12} take on fixed values, $\{k, k_{12}\} = \{0.6, 1.0\}$ N/m, but these values are not known (i.e. epistemic uncertainty). In order to infer k and k_{12}, measurements are obtained in the form of the eigenfrequencies $\boldsymbol{D} = \{\omega_1, \omega_2\}$ whereby ω_1 and ω_2 are the 2 eigenfrequencies associated with the in-phase and out-of-phase mode shapes of the symmetric system respectively which can be easily computed as follows [68]:

$$\hat{\omega}_1 = \sqrt{\frac{k}{m}} \tag{23}$$

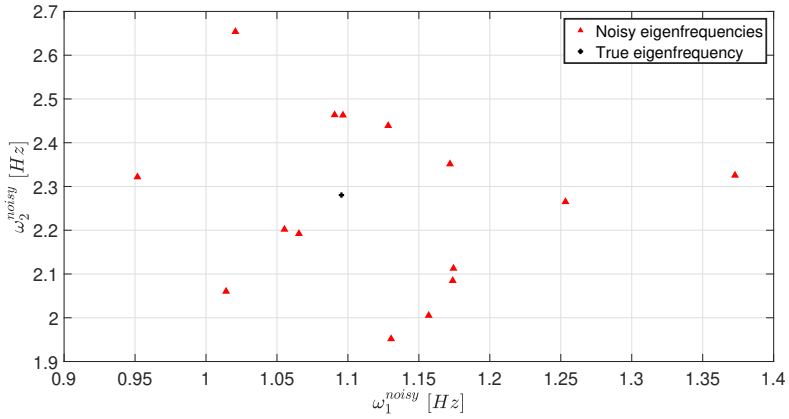


Figure 4: Scatterplot of the 15 different "measured" values ω_1 and ω_2.

$$\hat{\omega}_2 = \sqrt{\frac{(k + 2 \cdot k_{12})}{m}} \tag{24}$$

The above equations constitute the model class $M = \{\hat{\omega}_1, \hat{\omega}_2\}$ to be updated. The frequency measurements of ω_1 and ω_2 are however corrupted with "noise" ϵ_1 and ϵ_2 respectively such that:

$$\omega_1 = \hat{\omega}_1 + \epsilon_1 \tag{25}$$

$$\omega_2 = \hat{\omega}_2 + \epsilon_2 \tag{26}$$

whereby ϵ_1 and ϵ_2 are the "noise" terms following a Normal distribution with means 0 Hz and standard deviations σ_1 and σ_2 respectively. Here, σ_1 and σ_2 are fixed values set at 10 % of the nominal values of ω_1 and ω_2 respectively. This yields $\{\sigma_1, \sigma_2\} = \{0.110, 0.228\}\ Hz$. For simplicity, it is assumed that the measurement "noise" ϵ_1 and ϵ_2 are not correlated and that the "noise" between individual measurements of ω_1 and ω_2 are also independent. In this problem, 15 independent realizations of ω_1 and ω_2 are obtained and these synthetic data are presented in the form of a scatterplot shown in Figure 4 while the numerical values are presented in Table 1.

Measurement No.	ω_1 $[Hz]$	ω_2 $[Hz]$	Measurement No.	ω_1 $[Hz]$	ω_2 $[Hz]$
1	1.172	2.351	9	1.055	2.202
2	1.097	2.463	10	1.253	2.265
3	1.157	2.005	11	0.952	2.322
4	1.091	2.464	12	1.130	1.952
5	1.021	2.654	13	1.174	2.085
6	1.373	2.325	14	1.066	2.192
7	1.174	2.113	15	1.014	2.060
8	1.128	2.439	–	–	–

Table 1: Numerical values of ω_1 and ω_2 shown in Figure 4.

4.1.1. Bayesian Model Updating set-up

For this problem, the priors for k and k_{12} are set to be Uniform priors taking values between 0.01 N/m and 4.0 N/m. In addition, despite σ_1 and σ_2 being predetermined values, in reality, these 2 parameters are unknown and are also set as epistemic parameters to be inferred. The priors for σ_1 and σ_2 are also set to be Uniform priors taking values between 1.0×10^{-5} Hz and 1.0 Hz. Therefore, the total number of epistemic parameters to 4, thereby making this a 4-dimensional Bayesian model updating problem: $\boldsymbol{\theta} = \{k, k_{12}, \sigma_1, \sigma_2\}$. It is assumed that the epistemic parameters are independent from one another. The likelihood function is modelled to follow a Normal distribution and assuming independence between individual observations, it is expressed as follows:

$$P(\boldsymbol{D}|\boldsymbol{\theta}, M) = \prod_{n=1}^{15} \frac{1}{2 \cdot \pi \cdot \sigma_1 \cdot \sigma_2} \cdot exp\left[-\frac{(\omega_{1,n} - \hat{\omega}_1)^2}{2 \cdot {\sigma_1}^2} - \frac{(\omega_{2,n} - \hat{\omega}_2)^2}{2 \cdot {\sigma_2}^2}\right] \quad (27)$$

4.1.2. Results

From the posterior $P(\boldsymbol{\theta}|\boldsymbol{D}, M)$, 1000 samples are generated using the TMCMC sampler and the proposed TEMCMC sampler and the sampling time elapsed is of 48.60 s and 6.04 s respectively. Both samplers required 8 iterations . Figure 5 shows the evolution of the statistics of β_j and of the acceptance rates across all iterations j for both TMCMC and TEMCMC samplers. From

these plots, it can be seen that the evolution of β_j across all j are identical for both the TMCMC and TEMCMC samplers while the acceptance rates for the TMCMC sampler shows a higher degree of variation compared to that of the TEMCMC sampler. In addition, it can also be observed that out of 8 iterations, only 3 lie within the optimal limits of acceptance rate for the case of the TMCMC sampler and 7 for the case of the TEMCMC sampler. This highlights the effectiveness of the proposed adaptive tuning algorithm in moderating the step-size u in the TEMCMC sampler such that for majority of the iterations, the acceptance rates of the sampler is kept within the optimal range.

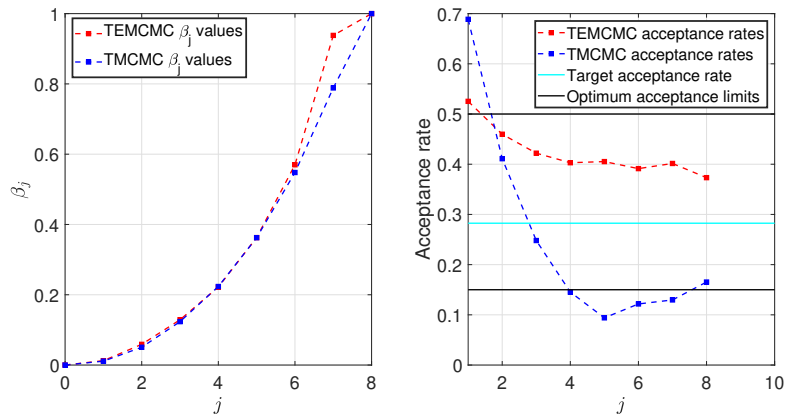


Figure 5: The statistics of β_j and the acceptance rates across all iterations j. The target acceptance rate is 0.283.

The resulting scatterplot matrix of the posterior samples is shown in Figure 6 and from there, the numerical results of the estimates of the epistemic parameters by the respective samplers are obtained. These results are presented and summarized in Table 2. From these results, it is possible to conclude that the TEMCMC sampler is generally able to perform as well as the TMCMC sampler and does so with a relatively lower computational cost. This highlights the relative computational efficiency of the TEMCMC sampler.

Finally, the resulting scatterplot of the model output is compared against the scatterplot of the data for ω_1 and ω_2. The results of the model updating

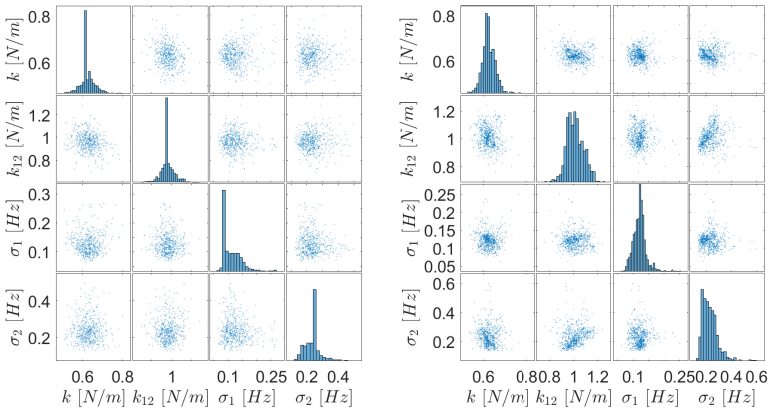


Figure 6: Scatterplot matrix illustrating the resulting posterior of the epistemic parameters.

Parameter, $\boldsymbol{\theta}$	Reference value	TMCMC		TEMCMC	
		Mean, $E[\boldsymbol{\theta}]$	COV [%]	Mean, $E[\boldsymbol{\theta}]$	COV [%]
k	0.6 N/m	0.625 N/m	5.67	0.625 N/m	5.67
k_{12}	1.0 N/m	0.962 N/m	6.48	1.013 N/m	6.80
σ_1	0.110 Hz	0.114 Hz	28.29	0.121 Hz	17.25
σ_2	0.228 Hz	0.236 Hz	20.86	0.229 Hz	26.15

Table 2: A summary of the statistics of the estimation of the epistemic parameters $\boldsymbol{\theta} = \{k, k_{12}, \sigma_1, \sigma_2\}$ via the posterior samples obtained using the TMCMC and TEMCMC samplers.

for the case of the TMCMC sampler and the TEMCMC sampler are illustrated in Figure 7. As seen from the figure, the resulting model outputs from the posterior samples obtained encompass the true solution of ω_1 and ω_2 for both the TMCMC and TEMCMC samplers. This highlights that the TEMCMC sampler works just as effectively as the TMCMC sampler as a tool in Bayesian model updating.

4.2. 2-D Multi-modal Posterior

The following set-up is based on the problem presented in [70] and the objective of this numerical example is to observe and compare the differences in

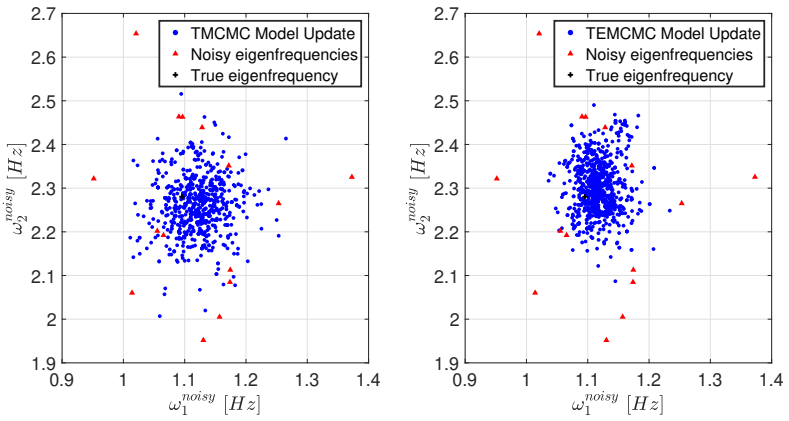


Figure 7: Scatterplot matrix illustrating the resulting posterior of the epistemic parameters.

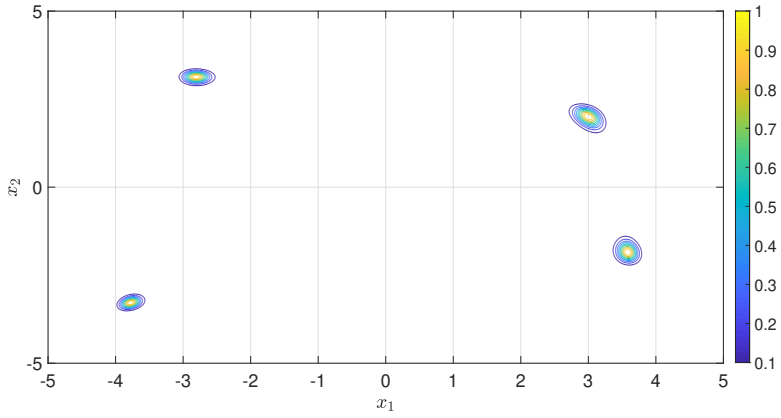


Figure 8: Contour plot illustration of the 4-peaked posterior based on the Himmelblau's function [69]. The numbers on the colour chart represent the height of the posterior computed from Eq. (29).

the performance of the TMCMC sampler and the proposed TEMCMC sampler in generating samples from a multi-modal posterior. This comparison will be done on the basis of the sample distribution in relation to the analytical solution of the posterior, the computation time elapsed in sampling from the posterior, and the acceptance rates across the transition steps j.

In this study, a 2-dimensional posterior with 4 peaks, defined by the di-

mensionless variables x_1 and x_2, is presented and illustrated in Figure 8. The analytical function of this posterior is based upon the Himmelblau's function which is a test-function used in mathematical optimisation problems to test the performance of optimisation algorithms. The Himmelblau's function $H(x_1, x_2)$ is mathematically defined as [69]:

$$H(x_1, x_2) = (x_1^2 + x_2 - 11)^2 + (x_1 + x_2^2 - 7)^2 \tag{28}$$

which yields 1 solution of local maximum at $\{x_1, x_2\} = \{-0.271, -0.923\}$ and 4 distinct solutions of local minima at $\{x_1, x_2\} = \{3.0, 2.0\}$, $\{-2.805, 3.131\}$, $\{-3.779, -3.283\}$, and $\{3.584, -1.848\}$. From which, the posterior of interest is then defined as follows [70]:

$$P(\boldsymbol{\theta}|\boldsymbol{D}, M) \propto exp\left[-H(x_1, x_2)\right] \tag{29}$$

which ensures that the local minimia of $H(x_1, x_2)$ now becomes the region of high probability giving rise to the 4 peaks.

4.2.1. Bayesian Model Updating set-up

For this problem, the epistemic parameters are x_1 and x_2, thereby making this a 2-dimensional Bayesian model updating problem: $\boldsymbol{\theta} = \{x_1, x_2\}$. The priors for x_1 and x_2 are set to be Uniform priors taking values between -5 and 5 and it is assumed that the epistemic parameters are independent from one another. The likelihood function is modelled as the exponential function of $-H(x_1, x_2)$ and thus takes on the same mathematical form as the posterior in Eq. (29).

4.2.2. Results

From the posterior $P(\boldsymbol{\theta}|\boldsymbol{D}, M)$, 1000 samples are generated using the TMCMC sampler and the proposed TEMCMC sampler and the sampling time elapsed is of 16.14 s and 4.96 s respectively. Both samplers required 5 iterations . Figure 9 shows the evolution of the statistics of β_j and of the acceptance rates across all iterations j for both TMCMC and TEMCMC samplers. From these

plots, it can be seen that the evolution of β_j across all j are identical for both the TMCMC and TEMCMC samplers while the acceptance rates for the TMCMC sampler shows a higher rate of decrease compared to that of the TEMCMC sampler, especially between $j = 1$ and $j = 2$. In addition, it can also be observed that out of 5 iterations, only 2 lie within the optimal limits of acceptance rate for the case of the TMCMC sampler and 3 for the case of the TEMCMC sampler. Futhermore, it can be seen that the acceptance rates for TEMCMC appears to tend towards the target acceptance rate value with increasing j while this is not the case for the TMCMC sampler. This indicates the effectiveness of the adaptive tuning algorithm for step-size u in the TEMCMC sampler in ensuring the majority of the iterations have acceptance rates within the optimal limits and that the tuning is done with reference to the target acceptance rate.

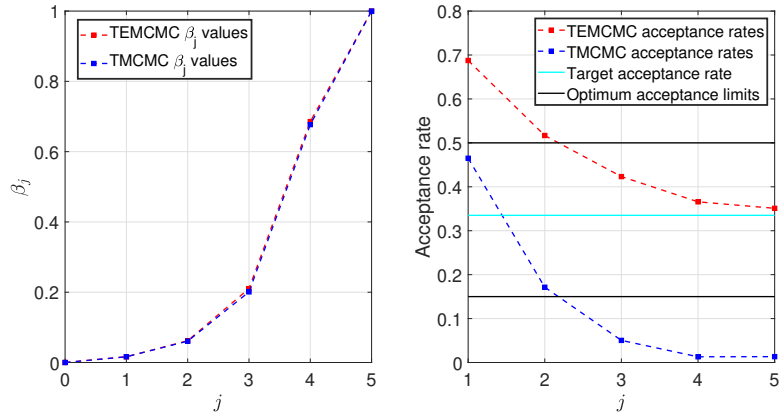


Figure 9: The statistics of β_j and the acceptance rates across all iterations j. The target acceptance rate is 0.335.

For the purpose of illustration, the scatterplots obtained from the transition distributions P^j between $j = 0$ and $j = 5$ via the TMCMC and TEMCMC samplers are presented in Figures 10 and 11 respectively. In both Figures 10 and 11, the scatterplots obtained by TMCMC and TEMCMC samplers are compared against the contour plot profile of the analytical solutions and it can be observed that in both figures, the scatterplot profiles match closely to the corresponding

contour plot profile. This validates the effectiveness of both samplers in sampling from all P^j. Figure 12 presents the scatterplot profile of the samples of the final posterior $P(\boldsymbol{\theta}|\boldsymbol{D}, M)$ in comparison with the contour plot profile of the analytical solution. While both samplers are able to sample effectively from the 4-peaked posterior, upon closer inspection, it can be observed that there is significantly less exploration of the sample space by the samples obtained via the TMCMC sampler compared to the TEMCMC sampler. This is due to the TMCMC sampler having a very low rate of acceptance in the latter iterations $j = 4$ and $j = 5$ where the acceptance rates are approximately 0.013 as seen in Figure 9. This reduces the number of unique samples generated from the MCMC step of the TMCMC algorithm leading to many repeated samples and a poor exploration of the sample space.

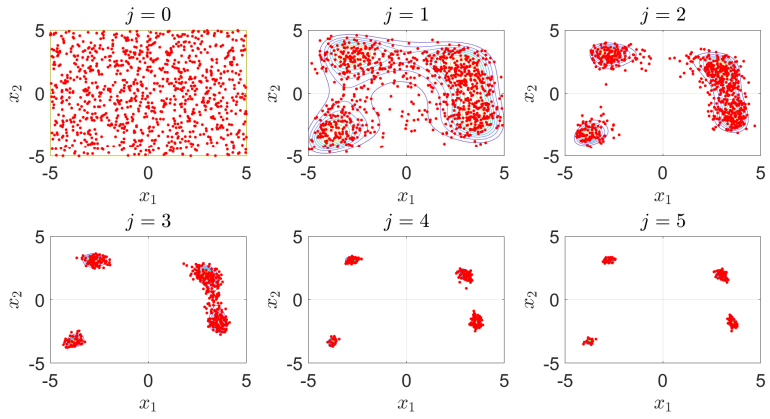


Figure 10: Scatterplots obtained from the transitional distributions P^j between $j = 0$ and $j = 5$ via TMCMC sampler. Each scatterplot is presented along with the contour plot profile of the analytical solution of P^j as comparison.

5. Application: Aluminium Frame Problem

The following example is based on the Aluminium Frame problem presented in [34]. As per illustrated in Figure 13, the structure consists of 7 beams (3 horizontals, 2 long verticals, and 2 short verticals) and 2 movable masses m_1

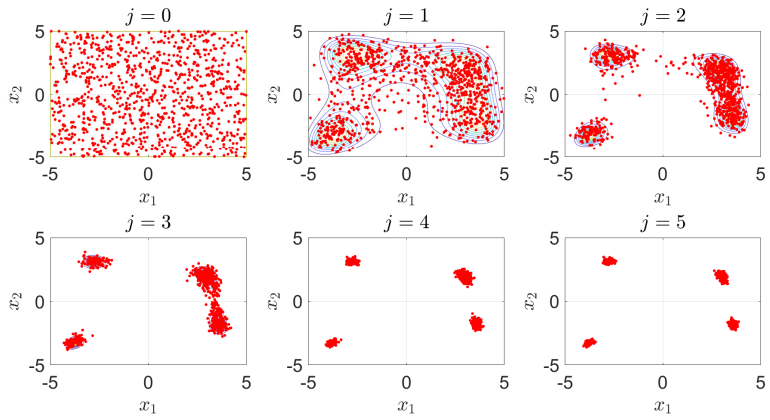


Figure 11: Scatterplots obtained from the transitional distributions P^j between $j = 0$ and $j = 5$ via TEMCMC sampler. Each scatterplot is presented along with the contour plot profile of the analytical solution of P^j as comparison.

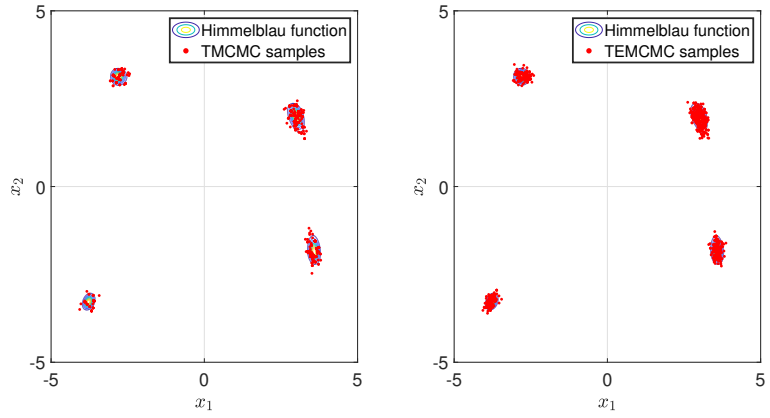


Figure 12: Resulting scatterplots of the samples from the final posterior $P(\boldsymbol{\theta}|\boldsymbol{D}, M)$ obtained via TMCMC (left) and TEMCMC (right) samplers along with the analytical contour plot profile as a comparison.

and m_2 with positions pm_1 and pm_2 respectively. The mass positions pm_1 and pm_2 are used to simulate structural damage.

Experimental data was collected and processed to determine for a given combination of $\{pm_1, pm_2\}$ the 6 natural frequencies corresponding to: ω_1 (1^{st} in-plane bending mode); ω_2 (1^{st} out-of-plane bending mode); ω_3 (1^{st} torsional

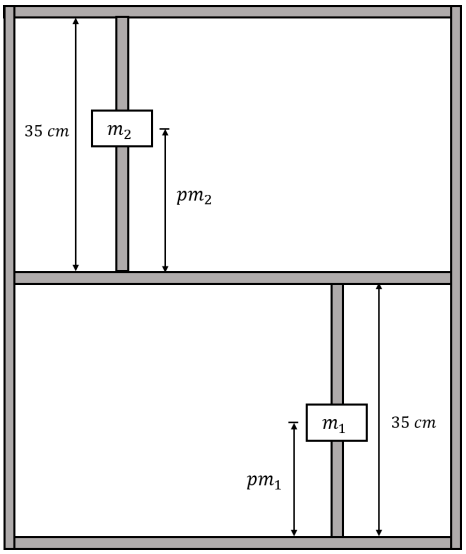


Figure 13: Schematic diagram of the Aluminium Frame with moveable masses m_1 and m_2 [34, 71, 72].

mode); ω_4 (2^{nd} in-plane bending mode); ω_5 (2^{nd} order out-of-plane bending mode); and ω_6 (2^{nd} torsional modes). A total of 11 sets of experimental data are generated from 11 distinct combinations of $\{pm_1, pm_2\}$ and these are summarised in Table 3:

5.1. Bayesian Model Updating set-up

For this problem, an Artificial Neural Network (ANN) is used as a surrogate model M in place of the computationally-expensive Finite Element Model (FEM) of the structure to perform Bayesian model updating and infer $\{pm_1, pm_2\}$. The architecture of the ANN comprises of 3 layers: 1 input-layer with 2 nodes, 1 hidden-layer with 10 nodes, and 1 output-layer with 6 nodes. For the purpose of calibrating the ANN, 103 simulated values of $\{pm_1, pm_2\}$ and $\{\omega_1, \ldots, \omega_6\}$ are obtained from the simulation database in [71]. The resulting 103 sets of simulated data are presented in the form of scatterplots as shown in Figures 14a and 14b.

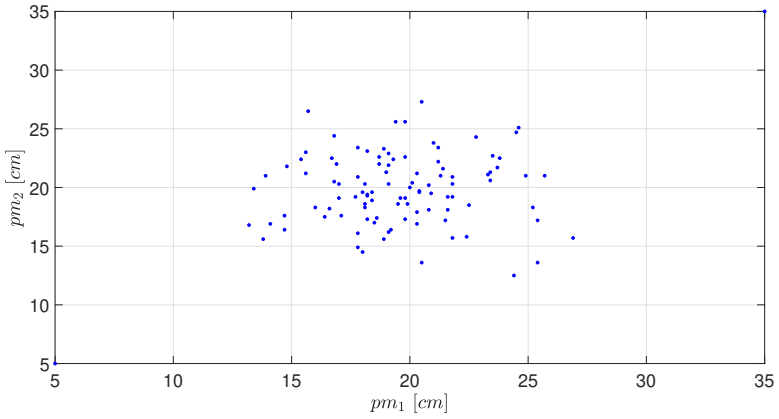
The calibration of the ANN was performed via the Feed-Forward Back-

Exp.	$\{pm_1, pm_2\}$ $[cm]$	ω_1 $[Hz]$	ω_2 $[Hz]$	ω_3 $[Hz]$	ω_4 $[Hz]$	ω_5 $[Hz]$	ω_6 $[Hz]$
1	$\{5, 5\}$	20.11	22.79	47.52	63.96	183.82	283.51
2	$\{5, 20\}$	18.72	20.46	46.97	72.24	214.84	296.32
3	$\{5, 35\}$	17.715	18.29	46.42	63.45	196.38	278.70
4	$\{20, 5\}$	19.40	22.39	46.32	61.78	173.49	259.76
5	$\{20, 20\}$	17.91	20.28	45.67	64.73	190.84	284.09
6	$\{20, 35\}$	16.71	18.21	45.18	56.53	177.97	264.44
7	$\{35, 5\}$	17.71	21.76	44.00	59.48	164.05	254.48
8	$\{35, 20\}$	16.91	19.82	43.15	60.06	175.75	279.10
9	$\{35, 35\}$	15.95	17.89	42.44	50.66	163.55	257.82
10	$\{11, 11\}$	19.58	21.73	47.00	67.54	196.21	285.95
11	$\{29, 29\}$	16.65	18.85	43.93	55.43	174.35	284.84

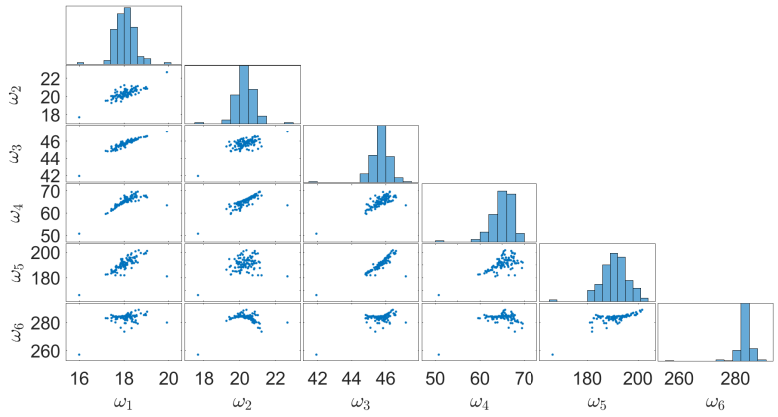
Table 3: A summary of the experimental data obtained from the hammer impact test. Data obtained from [71].

Propagation algorithm [73] with a sigmoidal activation function. Of the 103 sets of training data, 70 % was used to train the ANN, 15 % for validation, and 15 % for testing [34]. The calibration procedure took 0.513 s and yielded an overall regression coefficient R^2 of 0.9997 which indicates that the existing ANN architecture is sufficiently robust. A regression plot of the calibrated ANN is provided in Figure 15.

The Bayesian model updating set-up is as follows: The priors for pm_1 and pm_2 are set to be Uniform priors taking values between 5.0 cm and 35.0 cm. In addition, the measurement "noise" σ_v corresponding to the natural frequency ω_v (for $v = 1, \ldots, 6$) are also set to as epistemic parameters to be inferred. The prior for each σ_v is also set as a Uniform prior taking values between 0.001 Hz and 100.0 Hz. This brings the total number of epistemic parameters to 8, thereby making this a 8-dimensional Bayesian model updating problem: $\boldsymbol{\theta} = \{pm_1, pm_2, \sigma_1, \ldots, \sigma_6\}$. It is assumed that the epistemic parameters are independent from one another.



(a) Scatterplot of the 103 simulated values of $\{pm_1, pm_2\}$.



(b) Resulting scatterplot matrix of the response frequencies obtained from the FEM.

In [34], 3 likelihood functions are presented and used to perform Bayesian model updating. Defining M_v being the model output for ω_v, details to each of these likelihood functions are summarised in Table 4.

As shown in Table 4, likelihoods f_2 and f_3 are independent of σ_v. To ensure that the uncertainty of σ_v is accounted for, and that the influence of all the likelihood functions are also captured in the Bayesian model updating procedure, an approach would be to combine all 3 likelihoods. There are 3 ways to do so, under the assumption that likelihoods f_1, f_2, and f_3 have all equal influence, and the resulting combined likelihood functions with relevant references are

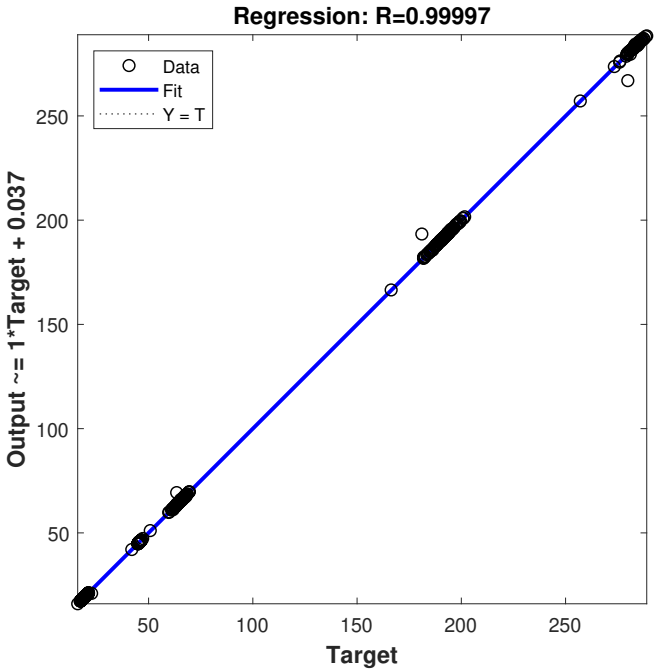


Figure 15: Regression plot of the calibrated ANN.

provided in Table 5.

Due to the uncertainty associated with the appropriate choice of combined likelihood function, the Robust Bayesian (RB) framework [80, 81] is adopted here. Using this framework, 3 distinct posteriors are derived by combining the prior with each of the combined likelihoods. Let the m^{th} posterior $P_m(\boldsymbol{\theta}|\boldsymbol{D}, M)$ be defined as the product of the prior and combined likelihood L_m (for $m = 1, \ldots, 3$). From there, a P-box [82, 83] can be constructed from the Empirical Cumulative Distribution Function (ECDF) of the posteriors where the 95 % Credible Interval (CI) for pm_1 and pm_2 are obtained. Such analysis is performed for all 11 sets of experiments from the data presented in Table 3.

Symbol	Type	Likelihood function, $P(\boldsymbol{D}\|\boldsymbol{\theta}, M)$
f_1	Normal Distribution	$\prod_{v=1}^{6} \frac{1}{\sigma_v \cdot \sqrt{2\pi}} \cdot exp\left[-\frac{(\omega_v - M_v)^2}{2 \cdot {\sigma_v}^2}\right]$
f_2	Inverse Squared Error	$\prod_{v=1}^{6} 1 - exp\left[-\frac{1}{(\omega_v - M_v)^2}\right]$
f_3	Inverse Error	$\prod_{v=1}^{6} 1 - exp\left[-\sqrt{\frac{1}{(\omega_v - M_v)^2}}\right]$

Table 4: The likelihood functions employed in [34] for Bayesian inference of pm_1 and pm_2.

5.2. Results

From $P_m(\boldsymbol{\theta}|\boldsymbol{D}, M)$, 1000 samples are generated using the TEMCMC sampler. The P-boxes obtained for each experiment are presented in Figure 16 for the case of pm_1, and Figure 17 for the case of pm_2. The numerical results of the 95 % CI obtained from the P-boxes for pm_1 and pm_2 for each experiment are summarised in Table 6. As shown in Table 6, the 95 % CI obtained for pm_1 generally encompasses the true value for most experiments with the exception of experiments where the true values of pm_1 are 5 cm and 35 cm. This is also observed for the case of the 95 % CI obtained for pm_2. A key reason for this is the poor performance of the ANN in regions of the sample space close to these values. To substantiate this, it can be observed from Figure 14a that most of the input samples used to train the ANN are situated in the interval $[15, 25]$ cm in both pm_1 and pm_2 coordinates and are sparsely distributed outside this region. In addition, it can also be seen from Figure 15 that the data is not evenly distributed along the regression plot which implies that the ANN does not mimic the general behaviour displayed by the FE model in regions where

Symbol	Type	Combined Likelihood Function	Reference(s)
L_1	Weighted Addition	$\frac{1}{3}\sum_{l=1}^{3} f_l$	[74, 75]
L_2	Bayes' Multiplication	$\prod_{l=1}^{3} f_l$	[74, 76, 77, 78]
L_3	Aggregated Function	$\sqrt{\frac{1}{3}\sum_{l=1}^{3} f_l^2}$	[74, 79]

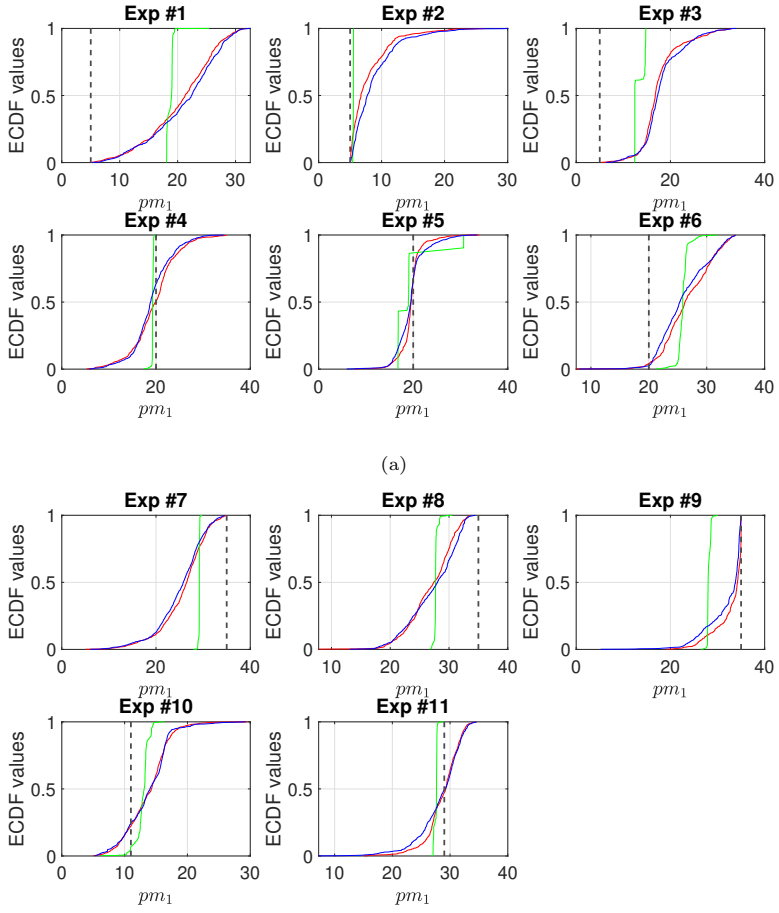
Table 5: The different forms of combined likelihood functions.

training data is not obtained. This has been highlighted in [34].

The resulting sampling time elapsed, number of iterations, and range of acceptance rates across iterations are summarised in Table 7. As seen from the table, the acceptance rate interval by the TEMCMC sampler generally encompasses the optimal range of values for all experiments and choice of combined likelihood functions. This highlights the robustness of the proposed adaptive tuning algorithm in moderating the acceptance rates of the sampler. In addition, it can also be seen that the number of iterations required by the TEMCMC sampler to generate samples from $P_2(\boldsymbol{\theta}|\boldsymbol{D}, M)$ is the highest while a similar number of iterations was required for $P_1(\boldsymbol{\theta}|\boldsymbol{D}, M)$ and $P_3(\boldsymbol{\theta}|\boldsymbol{D}, M)$.

To further investigate this, the influence of the individual likelihoods within L_1, L_2, and L_3. For each L_m, the samples $\boldsymbol{\theta}_i$ from $P_m(\boldsymbol{\theta}|\boldsymbol{D}, M)$ in a given experiment are used to compute the likelihood values of f_1, f_2, and f_3. These likelihood values are normalised to obtain the relative influence of f_l, denoted by Γ_{f_l}:

$$\Gamma_{f_l} = \frac{f_l}{\sum_{l=1}^{3} f_l} \tag{30}$$



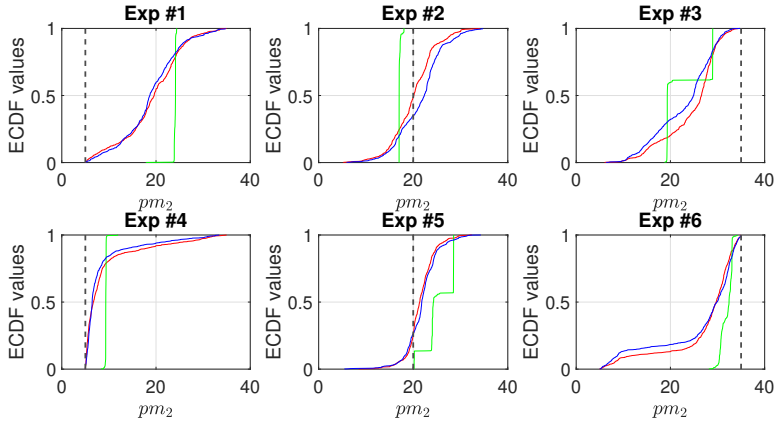
(a)

(b)

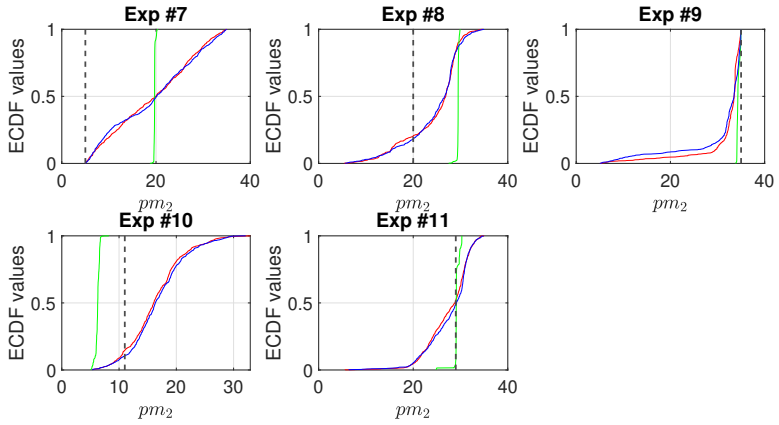
Figure 16: P-boxes for pm_1 for the respective experiments constructed from the ECDFs of $P_1(\boldsymbol{\theta}|\boldsymbol{D}, M)$ (red), $P_2(\boldsymbol{\theta}|\boldsymbol{D}, M)$ (green), and $P_3(\boldsymbol{\theta}|\boldsymbol{D}, M)$ (blue). The black dotted vertical line denotes the true value.

From there, the mean and COV of Γ_{f_l} is computed. This procedure is repeated for all experiments. The statistics of Γ_{f_l} for L_1, L_2, and L_3 across all experiments are presented in Tables 8, 9, and 10 respectively.

From Tables 8 to 10, it can be observed that the value of Γ_{f_2} is consistently the highest across all experiments. Given that the analytical forms of L_1 and L_3 involve taking the sum of the individual and the square of each likelihood



(a)



(b)

Figure 17: P-boxes for pm_2 for the respective experiments constructed from the ECDFs of $P_1(\boldsymbol{\theta}|\boldsymbol{D}, M)$ (red), $P_2(\boldsymbol{\theta}|\boldsymbol{D}, M)$ (green), and $P_3(\boldsymbol{\theta}|\boldsymbol{D}, M)$ (blue). The black dotted vertical line denotes the true value.

respectively, the overall likelihood values for L_1 and L_3 would therefore be dominated by the value of f_2. On the other hand, Γ_{f_1} is consistently the smallest (i.e. close to 0) across all experiments within all 3 combined likelihood functions. Because the analytical form of L_2 is essentially a direct product of all 3 likelihood functions, the value of f_1 would cause the value of L_2 to be significantly smaller compared to L_1 and L_3. As a result , the value of $\Delta\beta_j$ computed by the

Exp.	True value pm_1 $[cm]$	95 % CI pm_1 $[cm]$	True value pm_2 $[cm]$	95 % CI pm_2 $[cm]$
1	5	$[7.56, 30.23]$	5	$[5.65, 32.04]$
2	5	$[5.07, 19.03]$	20	$[10.26, 31.22]$
3	5	$[9.83, 29.34]$	35	$[10.70, 32.45]$
4	20	$[9.13, 28.83]$	5	$[5.06, 30.59]$
5	20	$[14.68, 30.67]$	20	$[15.16, 29.58]$
6	20	$[19.38, 34.01]$	35	$[6.02, 34.51]$
7	35	$[12.86, 33.69]$	5	$[5.55, 34.14]$
8	35	$[18.53, 32.92]$	20	$[9.51, 32.35]$
9	35	$[22.71, 34.99]$	35	$[8.07, 34.97]$
10	11	$[6.54, 20.87]$	11	$[5.31, 27.97]$
11	29	$[18.28, 33.18]$	29	$[18.78, 33.66]$

Table 6: Results of the 95 % CI obtained from the P-boxes for pm_1 and pm_2 for each experiment.

TEMCMC algorithm will be smaller for the case of L_2, therefore, requiring a larger number of iterations for obtaining the posterior $P_2(\boldsymbol{\theta}|\boldsymbol{D}, M)$. Similarly, when L_1 and L_3 take higher values, the value of $\Delta\beta_j$ computed by the TEMCMC algorithm is larger, resulting in less iterations.

6. Conclusions

An efficient and robust sampler named Transitional Ensemble Markov Chain Monte Carlo has been proposed for Bayesian inference. The proposed sampler uses an affine-invariant ensemble sampler in place of the traditional Metropolis-Hasting sampler and includes an adaptive tuning algorithm making the approach "tune-free" for users. The proposed sampler out-performs the current samplers available in sampling from highly-skewed, anisotropic distributions [32, 66] such as the transition distributions by exploiting the advantages of the Transitional Markov Chain Monte Carlo . The absence of a proposal distribution reduces the number of parameters to tune adaptively, thereby reducing the computational

Exp.	$P_1(\boldsymbol{\theta}\|\boldsymbol{D}, M)$			$P_2(\boldsymbol{\theta}\|\boldsymbol{D}, M)$			$P_3(\boldsymbol{\theta}\|\boldsymbol{D}, M)$		
	Time $[s]$	Iterations	Acceptance	Time $[s]$	Iterations	Acceptance	Time $[s]$	Iterations	Acceptance
1	124.88	2	$[0.452, 0.478]$	687.63	11	$[0.332, 0.474]$	126.99	2	$[0.438, 0.497]$
2	178.40	3	$[0.421, 0.503]$	758.29	12	$[0.342, 0.477]$	172.20	3	$[0.426, 0.488]$
3	127.22	2	$[0.441, 0.498]$	634.53	10	$[0.353, 0.491]$	123.09	2	$[0.450, 0.488]$
4	177.76	3	$[0.429, 0.490]$	526.76	8	$[0.374, 0.545]$	169.55	3	$[0.439, 0.494]$
5	181.36	3	$[0.425, 0.496]$	641.87	10	$[0.319, 0.476]$	177.73	3	$[0.424, 0.498]$
6	123.92	2	$[0.454, 0.500]$	758.80	11	$[0.344, 0.489]$	119.62	2	$[0.435, 0.490]$
7	144.16	2	$[0.456, 0.506]$	757.30	10	$[0.358, 0.473]$	126.88	2	$[0.452, 0.494]$
8	129.26	2	$[0.431, 0.500]$	669.17	12	$[0.376, 0.494]$	121.28	2	$[0.433, 0.511]$
9	242.37	4	$[0.414, 0.488]$	665.24	9	$[0.360, 0.476]$	179.12	3	$[0.419, 0.493]$
10	206.06	3	$[0.431, 0.497]$	728.62	11	$[0.352, 0.486]$	169.32	3	$[0.435, 0.494]$
11	168.19	3	$[0.436, 0.506]$	650.88	11	$[0.369, 0.494]$	168.41	3	$[0.429, 0.488]$

Table 7: Summary of the sampling time elapsed, number of iterations, and range of acceptance rates across iterations by the TEMCMC sampler for the respective posteriors in each experiment.

Exp.	f_1		f_2		f_3	
	Mean Γ_{f_1}	COV [%]	Mean Γ_{f_2}	COV [%]	Mean Γ_{f_3}	COV [%]
1	4.84×10^{-53}	3.13×10^{3}	0.979	2.18	0.021	100.42
2	1.38×10^{-51}	2.92×10^{3}	0.943	4.87	0.058	79.87
3	5.30×10^{-52}	2.75×10^{3}	0.931	6.08	0.069	82.07
4	4.21×10^{-53}	1.88×10^{3}	0.975	2.24	0.026	85.83
5	2.47×10^{-46}	2.12×10^{3}	0.583	29.18	0.417	40.83
6	2.03×10^{-51}	2.47×10^{3}	0.954	3.65	0.046	76.39
7	2.23×10^{-52}	2.19×10^{3}	0.999	0.07	0.001	69.49
8	2.55×10^{-52}	1.66×10^{3}	0.904	8.54	0.096	80.60
9	3.62×10^{-48}	2.81×10^{3}	0.920	11.07	0.080	126.97
10	2.13×10^{-52}	2.81×10^{3}	0.810	18.89	0.189	80.97
11	5.58×10^{-54}	2.20×10^{3}	0.931	3.16	0.069	42.54

Table 8: Statistics of f_1, f_2, and f_3 for combined likelihood L_1.

Exp.	f_1		f_2		f_3	
	Mean Γ_{f_1}	COV [%]	Mean Γ_{f_2}	COV [%]	Mean Γ_{f_3}	COV [%]
1	1.95×10^{-23}	3.15×10^3	0.985	0.30	0.015	19.34
2	1.74×10^{-22}	9.26×10^2	0.920	0.84	0.080	9.63
3	1.57×10^{-30}	3.33×10^2	0.968	3.84	0.032	116.57
4	6.01×10^{-37}	1.53×10^3	0.996	0.03	0.004	6.24
5	2.54×10^{-25}	6.35×10^2	0.690	30.60	0.311	67.96
6	1.90×10^{-13}	3.11×10^3	0.960	1.53	0.041	36.08
7	9.12×10^{-30}	3.04×10^2	0.999	0.002	0.001	1.40
8	3.06×10^{-24}	1.20×10^3	0.879	2.51	0.121	18.30
9	2.04×10^{-28}	7.93×10^2	0.994	0.10	0.006	17.43
10	1.68×10^{-10}	1.99×10^3	0.995	0.17	0.005	35.95
11	1.44×10^{-20}	2.93×10^3	0.933	1.63	0.067	22.69

Table 9: Statistics of f_1, f_2, and f_3 for combined likelihood L_2.

Exp.	f_1		f_2		f_3	
	Mean Γ_{f_1}	COV [%]	Mean Γ_{f_2}	COV [%]	Mean Γ_{f_3}	COV [%]
1	5.51×10^{-52}	2.64×10^3	0.978	2.34	0.022	102.67
2	6.91×10^{-54}	2.55×10^3	0.945	5.09	0.055	87.13
3	4.60×10^{-48}	2.23×10^3	0.944	5.45	0.056	91.15
4	5.41×10^{-47}	3.12×10^3	0.973	1.98	0.027	72.24
5	1.24×10^{-49}	2.33×10^3	0.638	28.78	0.362	50.77
6	2.97×10^{-55}	2.44×10^3	0.954	3.56	0.046	73.20
7	2.37×10^{-55}	2.77×10^3	0.999	0.07	0.001	65.54
8	7.54×10^{-51}	3.03×10^3	0.914	8.19	0.086	87.34
9	1.05×10^{-51}	3.16×10^3	0.952	7.41	0.048	146.69
10	1.46×10^{-53}	1.94×10^3	0.802	18.77	0.198	76.19
11	4.42×10^{-54}	2.23×10^3	0.932	3.33	0.068	45.83

Table 10: Statistics of f_1, f_2, and f_3 for combined likelihood L_3.

cost and eradicating the model uncertainty in considering the choice of the distribution model.

To illustrate the efficiency and effectiveness of the proposed Transitional Ensemble Markov Chain Monte Carlo sampler, two numerical examples and one experimental example have been presented showing the applicability of the approach for performing Bayesian model updating also in the presence of a complex a multi-modal posterior. In all the examples investigated, the Transitional Ensemble Markov Chain Monte Carlo sampler outperforms the traditional sampler by requiring a shorter time whilst ensuring that the acceptance rates are well-moderated within optimal bounds. Furthermore, through the experimental example, we have also demonstrated the robustness of the proposed sampler in performing Bayesian inference using measurements obtained under realistic settings. This highlights the applicability of the Transitional Ensemble Markov Chain Monte Carlo in addressing real-world engineering problems.

In summary, the results from the examples presented in this paper highlights the following key strengths of the proposed Transitional Ensemble Markov Chain Monte Carlo sampler: 1) it allows for the sampling of badly-scaled and highly-anisotropic distributions without requiring extra computational costs; 2) it is free from tuning by the user; and 3) it is more robust than the Transitional Markov Chain Monte Carlo in controlling the acceptance rates automatically. One significant drawback of this sampler, however, is the relative complexity in coding the Transitional Ensemble Markov Chain Monte Carlo algorithm from scratch compared to the standard Transitional Markov Chain Monte Carlo. To address this issue, access to the MATLAB code is provided on GitHub: https://github.com/Adolphus8/Transitional_Ensemble_MCMC.git

7. Appendix

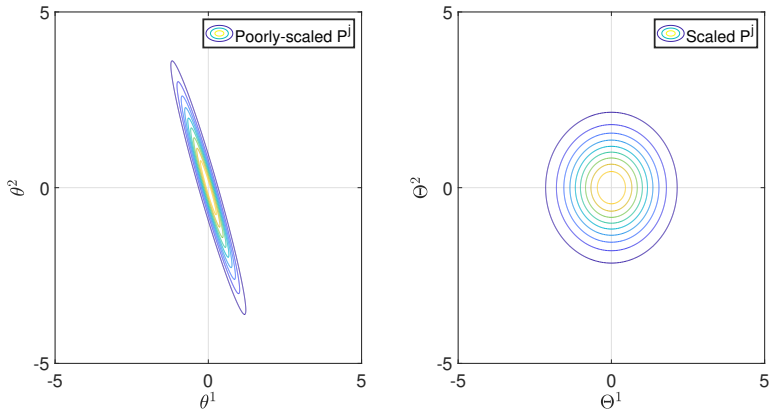


Figure 18: Contour plots illustrating the skewed P^j defined by Eq. (31) (left) and the scaled, isotropic P^j in the affine-transformed space defined by Eq. (33) (right).

In this illustrative example, a skewed "transition" distribution defined in a $2D$ sample space $\boldsymbol{\theta} = \{\theta^1, \theta^2\}$ is presented in Figure 18 with the following mathematical expression:

$$P^j(\boldsymbol{\theta}) \propto \left\{ exp\left[-\frac{(3 \cdot \theta^1 + \theta^2)^2}{0.08} - \frac{(\theta^1 - \theta^2)^2}{2} \right] \right\}^{\beta_j} \tag{31}$$

where $\beta_j = 0.2$. To simplify the distribution such that it becomes easier to generate samples from, one could re-scale the problem via the following affine-transformation:

$$\begin{bmatrix} \Theta^1 \\ \Theta^2 \end{bmatrix} = \begin{bmatrix} 15 \cdot \sqrt{\beta_j} & 5 \cdot \sqrt{\beta_j} \\ \sqrt{\beta_j} & -\sqrt{\beta_j} \end{bmatrix} \begin{bmatrix} \theta^1 \\ \theta^2 \end{bmatrix} \tag{32}$$

This yields a relatively simpler isotropic distribution $P'^j(\boldsymbol{\Theta})$:

$$P'^j(\boldsymbol{\Theta}) \propto exp\left[-\frac{(\Theta^1)^2}{2} - \frac{(\Theta^2)^2}{2} \right] \tag{33}$$

From both $P^j(\boldsymbol{\theta})$ and $P'^j(\boldsymbol{\Theta})$, 1000 samples are obtained across 4 chains (i.e. 250 samples per chain) using AIES and MH samplers. The tuning-parameter

settings for the respective samplers are presented in Table 11 which ensures that the acceptance rates for both samples are as close to 0.234 as possible [46]. Following which, the samples in the $\boldsymbol{\Theta}$-space obtained by the respective samplers would be re-scaled to the $\boldsymbol{\theta}$-space via the inverse of Eq. (32). This yields the results illustrated in Figure 19 where it can be seen that the ECDF obtained directly from $P^j(\boldsymbol{\theta})$ and that re-scaled from $P'^j(\boldsymbol{\Theta})$ are in very good agreement. This is quantified by the area enclosed by both ECDFs where it can be seen from Table 12 that this area is small (i.e. close to 0) compared to that for the case of the MH sampler. In addition, it can also be observed from Figure 19 that the profile of the ECDF obtained directly from $P^j(\boldsymbol{\theta})$ by MH sampler (i.e. in purple) deviates significantly from that of the analytical CDF of θ^1 and θ^2. In the case of the AIES, such deviation is less significant in both dimensions. These results highlight not only the capability of the AIES in sampling directly from a skewed distribution without the need to re-scale such distribution under an affine-transformation, but also its affine-invariant property which allows it to sample from such distribution as effectively as it would from a scaled isotropic distribution. These characteristics are not exhibited by the MH sampler.

Tuning-parameter	Case: $P^j(\boldsymbol{\theta})$	Case: $P'^j(\boldsymbol{\Theta})$
Step-size, u (AIES)	8.0	8.0
Covariance matrix, $\boldsymbol{\Sigma}$ (MH)	$0.5 \cdot \boldsymbol{I}$	$5 \cdot \boldsymbol{I}$

Table 11: Parameter settings implemented for the respective samplers in sampling from $P^j(\boldsymbol{\theta})$ and $P'^j(\boldsymbol{\Theta})$ respectively. $\boldsymbol{I}$ denotes the identity matrix.

	AIES	MH
θ^1	0.049	0.265
θ^2	0.089	0.813

Table 12: Results of the area enclosed by the ECDF obtained directly from $P^j(\boldsymbol{\theta})$ and that re-scaled from $P'^j(\boldsymbol{\Theta})$ for the respective samplers.

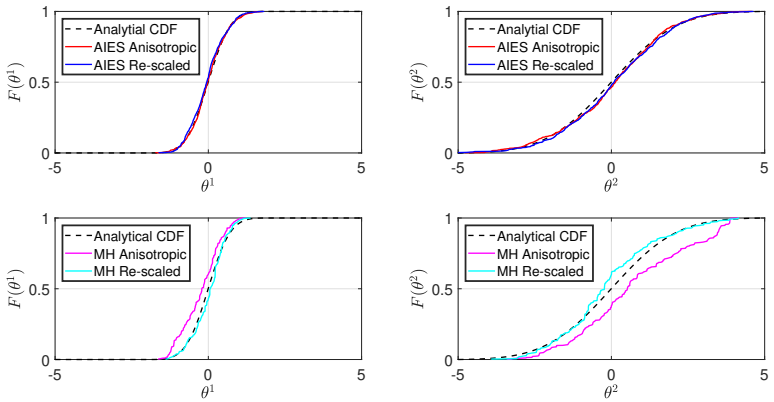


Figure 19: ECDFs of θ^1 and θ^2 obtained directly from $P^j(\boldsymbol{\theta})$ and those re-scaled from $P'^j(\boldsymbol{\Theta})$ when using MH and AIES.

References

[1] J. L. Beck, L. S. Katafygiotis, Updating Models and Their Uncertainties. I: Bayesian Statistical Framework, Journal of Engineering Mechanics 124 (1998) 455–461. doi:10.1061/(asce)0733-9399(1998)124:4(455).

[2] L. S. Katafygiotis, J. L. Beck, Updating Models and Their Uncertainties. II: Model Identifiability, Journal of Engineering Mechanics 124 (1998) 463––467. doi:10.1061/(asce)0733-9399(1998)124:4(463).

[3] H. Wan, W. Ren, Stochastic model updating utilizing Bayesian approach and Gaussian process model, Mechanical Systems and Signal Processing 70-71 (2016) 245–268. doi:10.1016/j.ymssp.2015.08.011.

[4] R. Astroza, N. Barrientos, Y. Li, E. I. Flores, Z. Liu, Bayesian updating of complex nonlinear FE models with high-dimensional parameter space using heterogeneous measurements and a batch-recursive approach, Engineering Structures 201 (2019) . doi:10.1016/j.engstruct.2019.109724.

[5] I. Behmanesh, B. Moaveni, Probabilistic identification of simulated damage on the Dowling hall footbridge through Bayesian finite element model

updating, Structural Control and Health Monitoring 22 (2014) 463–483. doi:10.1002/stc.1684.

[6] Y. Ni, Y. Wang, C. Zhang, A Bayesian approach for condition assessment and damage alarm of bridge expansion joints using long-term structural health monitoring data, Engineering Structures 212 (2020) . doi:10.1016/j.engstruct.2020.110520.

[7] S. Mishra, O. A. Vanli, B. P. Alduse, S. Jung, Hurricane loss estimation in wood-frame buildings using Bayesian model updating: Assessing uncertainty in fragility and reliability analyses, Engineering Structures 135 (2017) 81–94. doi:10.1016/j.engstruct.2016.12.063.

[8] Y. Xin, H. Hao, J. Li, Z. Wang, H. Wan, W. Ren, Bayesian based nonlinear model updating using instantaneous characteristics of structural dynamic responses, Engineering Structures 183 (2019) 459–474. doi:10.1016/j.engstruct.2019.01.043.

[9] H. Lam, J. Hu, J. Yang, Bayesian operational modal analysis and Markov chain Monte Carlo-based model updating of a factory building, Engineering Structures 132 (2017) 314–336. doi:10.1016/j.engstruct.2016.11.048.

[10] L. Ierimonti, I. Venanzi, N. Cavalagli, F. Comodini, F. Ubertini, An innovative continuous Bayesian model updating method for base-isolated RC buildings using vibration monitoring data, Mechanical Systems and Signal Processing 139 (2020) . doi:10.1016/j.ymssp.2019.106600.

[11] M. Song, B. Moaveni, C. Papadimitriou, A. Stavridis, Accounting for amplitude of excitation in model updating through a hierarchical Bayesian approach: Application to a two-story reinforced concrete building, Mechanical Systems and Signal Processing 123 (2019) 68–83. doi:10.1016/j.ymssp.2018.12.049.

[12] B. Goller, M. Broggi, A. Calvi, G. Schuëller, A stochastic model updating

technique for complex aerospace structures, Finite Elements in Analysis and Design 47 (2011) 739–752. doi:10.1016/j.finel.2011.02.005.

[13] J. Sun, J. Ellerbroek, J. M. Hoekstra, Aircraft initial mass estimation using Bayesian inference method, Transportation Research Part C: Emerging Technologies 90 (2018) 59–73. doi:10.1016/j.trc.2018.02.022.

[14] E. Patelli, Y. Govers, M. Broggi, H. M. Gomes, M. Link, J. E. Mottershead, Sensitivity or Bayesian model updating: A comparison of techniques using the DLR-AIRMOD test data, Archive of Applied Mechanics 87 (2017) 905–925. doi:10.1007/s00419-017-1233-1.

[15] A. Lye, A. Cicirello, E. Patelli, Sampling Methods for solving Bayesian Model Updating Problems: A Tutorial, Mechanical Systems and Signal Processing 159 (2021) 107760. doi:10.1016/j.ymssp.2021.107760.

[16] M. Pharr, W. Jakob, G. Humphreys, Monte Carlo Integration, Physically Based Rendering (2017) 747–802. doi:10.1016/b978-0-12-800645-0.50013-0.

[17] J. Y. Ching, Y. C. Chen, Transitional Markov Chain Monte Carlo Method for Bayesian Model Updating, Model Class Selection, and Model Averaging, Journal of Engineering Mechanics 133. doi:10.1061/(ASCE)0733-9399(2007)133:7(816).

[18] J. Y. Ching, J. S. Wang, Application of the Transitional Markov chain Monte Carlo algorithm to probabilistic site characterization, Engineering Geology 203 (2016) 151—167. doi:10.1016/j.enggeo.2015.10.015.

[19] G. A. Ortiz, D. A. Alvarez, D. Bedoya-Ruíz, Identification of Bouc–wen type models using the Transitional Markov Chain Monte Carlo method, Computers and Structures 146 (2015) 252–269. doi:10.1016/j.compstruc.2014.10.012.

[20] Y. Zhou, A. D. Popolo, Z. Chang, On the absence of a universal surface density, and a maximum Newtonian acceleration in dark matter haloes:

Consequences for MOND, Physics of the Dark Universe 28 (2020) . doi:10.1016/j.dark.2020.100468.

[21] J. Wang, L. Katafygiotis, Reliability-based optimal design of linear structures subjected to stochastic excitations, Structural Safety 47 (2014) 29––38. doi:10.1016/j.strusafe.2013.11.002.

[22] E. Patelli, M. Broggi, Y. Govers, J. E. Mottershead, Model Updating Strategy of the DLR-AIRMOD Test Structure, Procedia Engineering 199 (2017) 978–983. doi:10.1016/j.proeng.2017.09.221.

[23] E. Simoen, G. D. Roeck, G. Lombaert, Dealing with uncertainty in model updating for damage assessment: A review, Mechanical Systems and Signal Processing 56-57 (2015) 123–149. doi:10.1016/j.ymssp.2014.11.001.

[24] S. Sehgal, H. Kumar, Structural Dynamic Model Updating Techniques: A State of the Art Review, Archives of Computational Methods in Engineering 23 (2015) 515–533. doi:10.1007/s11831-015-9150-3.

[25] K. Yuen, Recent developments of Bayesian model class selection and applications in civil engineering, Structural Safety 32 (2010) 338–346. doi:10.1016/j.strusafe.2010.03.011.

[26] S. He, C. Ng, Guided wave-based identification of multiple cracks in beams using a Bayesian approach, Mechanical Systems and Signal Processing 84 (2017) 324–345. doi:10.1016/j.ymssp.2016.07.013.

[27] W. H. Zhou, F. Tan, K. V. Yuen, Model updating and uncertainty analysis for creep behavior of soft soil, Computers and Geotechnics 100 (2018) 135–143. doi:10.1016/j.compgeo.2018.04.006.

[28] P. Peralta, R. O. Ruiz, A. A. Taflanidis, Bayesian identification of electromechanical properties in piezoelectric energy harvesters, Mechanical Systems and Signal Processing 141 (2020) . doi:10.1016/j.ymssp.2019.106506.

[29] H. Jensen, D. Jerez, A stochastic framework for hydraulic performance assessment of complex water distribution networks: Application to connectivity detection problems, Probabilistic Engineering Mechanics 60 (2020) . doi:10.1016/j.probengmech.2020.103029.

[30] W. Betz, I. Papaioannou, D. Straub, Transitional Markov Chain Monte Carlo: Observations and Improvements, Journal of Engineering Mechanics 142 (2016) . doi:10.1061/(asce)em.1943-7889.0001066.

[31] G. O. Roberts, A. Gelman, W. R. Gilks, Weak Convergence and Optimal Scaling of Random Walk Metropolis Algorithms, The Annals of Applied Probability 7 (1997) 110–120. doi:10.1214/aoap/1034625254.

[32] J. Goodman, J. Weare, Ensemble samplers with affine invariance, Communications in Applied Mathematics and Computational Science 5 (2010) 65–80. doi:10.2140/camcos.2010.5.65.

[33] D. Foreman-Mackey, D. W. Hogg, D. Lang, J. Goodman, Emcee: The mcmc hammer, Publications of the Astronomical Society of the Pacific 125 (2013) 306–312. doi:10.1086/670067.

[34] R. Rocchetta, M. Broggi, Q. Huchet, E. Patelli, On-line Bayesian Model Updating for Structural Health Monitoring, Mechanical Systems and Signal Processing 103 (2018) 174–195. doi:10.1016/j.ymssp.2017.10.015.

[35] K. V. Yuen, *Bayesian Methods for Structural Dynamics and Civil Engineering*, Singapore: John Wiley and Sons Asia, 2010, ISBN: 978-0470824559.

[36] T. Bayes, Price, LII. An Essay towards Solving a problem in the Doctrine of Chances. By the late Rev. Mr. Bayes, F. R. S. communicated by Mr. Price, in a letter to John Canton, A. M. F. R. S., Philosophical Transactions of the Royal Society of London 53 (1763) 370–418. doi:10.1098/rstl.1763.0053.

[37] C. P. Robert, G. Casella, *Monte Carlo Statistical Methods*, 2nd Edition, Springer Science & Business Media, 2013, ISBN: 978-1475741452.

[38] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, D. B. Rubin, *Bayesian Data Analysis*, 3rd Edition, CRC Press, 2013, ISBN: 978-1439898208.

[39] W. K. Hastings, Monte Carlo Sampling Methods using Markov Chains and their Applications, Biometrika 57 (1970) 97—109. doi:10.1093/biomet/57.1.97.

[40] P. D. Moral, A. Doucet, A. Jasra, Sequential Monte Carlo Samplers, Journal of the Royal Statistical Society. Series B (Statistical Methodology) 68 (2006) 411–436.

[41] J. L. Beck, S. K. Au, Bayesian Updating of Structural Models and Reliability using Markov Chain Monte Carlo Simulation, Journal of Engineering Mechanics 128 (2002) 380–391. doi:10.1061/(ASCE)0733-9399(2002)128:4(380).

[42] P. Efraimidis, P. Spirakis, Weighted Random Sampling, Encyclopedia of Algorithms (2008) 1024–1027. doi:10.1007/978-0-387-30162-4_478.

[43] C. Andrieu, J. Thoms, A tutorial on adaptive MCMC, Statistics and Computing 18 (2008) 343–373. doi:10.1007/s11222-008-9110-y.

[44] L. Martino, V. Elvira, Metropolis sampling, Wiley StatsRef: Statistics Reference Online (2017) 1–18. doi:10.1002/9781118445112.stat07951.

[45] S. M. Lynch, *Introduction to applied Bayesian statistics and estimation for social scientists*, Springer Science and Business Media, 2007, ISBN: 978-0387712659.

[46] G. O. Roberts, J. S. Rosenthal, Optimal Scaling for various Metropolis-Hastings Algorithms, Statistical Science 16 (2001) 351–367. doi:10.1214/ss/1015346320.

[47] Y. Zhang, W. Yang, A comparative study of the stochastic simulation methods applied in structural health monitoring, Engineering Computations 31 (2014) 1484–1513. doi:10.1108/ec-07-2013-0185.

[48] A. Lye, A. Cicirello, E. Patelli, Bayesian Model Updating of Reliability Parameters using Transitional Markov Chain Monte Carlo with Slice Sampling, In Proceedings of the 30th European Safety and Reliability Conference and 15th Probabilistic Safety Assessment and Management Conference (2020) . doi:10.3850/978-981-14-8593-0_4374-cd.

[49] P. Angelikopoulos, C. Papadimitriou, P. Koumoutsakos, X-TMCMC: Adaptive kriging for Bayesian inverse modeling, Computer Methods in Applied Mechanics and Engineering 289 (2015) 409–428. doi:10.1016/j.cma.2015.01.015.

[50] Y. Jin, Z. Yin, W. Zhou, S. Horpibulsuk, Identifying parameters of Advanced Soil models using an enhanced transitional Markov chain Monte Carlo method, Acta Geotechnica 14 (2019) 1925–1947. doi:10.1007/s11440-019-00847-1.

[51] R. M. Neal, Slice Sampling, The Annals of Statistics 31 (2003) 705–767. doi:10.1214/aos/1056562461.

[52] F. Hou, J. Goodman, D. W. Hogg, J. Weare, C. Schwab, An Affine-Invariant Sampler For Exoplanet Fitting And Discovery In Radial Velocity Data, The Astrophysical Journal 745 (2012) 198. doi:10.1088/0004-637x/745/2/198.

[53] I. J. M. Crossfield, E. Petigura, J. E. Schlieder, A. W. Howard, B. J. Fulton, K. M. Aller, D. R. Ciardi, S. Lépine, T. Barclay, I. D. Pater, K. D. Kleer, E. V. Quintana, J. L. Christiansen, E. Schlafly, L. Kaltenegger, J. R. Crepp, T. Henning, C. Obermeier, N. Deacon, L. Weiss, H. T. Isaacson, B. M. S. Hansen, M. C. Liu, T. Greene, S. B. Howell, T. Barman, C. Mordasini, A Nearby M Star With Three Transiting Super-Earths Discovered By K2, The Astrophysical Journal 804 (2015) 10. doi:10.1088/0004-637x/804/1/10.

[54] A. Vanderburg, B. T. Montet, J. A. Johnson, L. A. Buchhave, L. Zeng, F. Pepe, A. C. Cameron, D. W. Latham, E. Molinari, S. Udry, C. Lovis,

J. M. Matthews, C. Cameron, N. Law, B. P. Bowler, R. Angus, C. Baranec, A. Bieryla, W. Boschin, D. Charbonneau, R. Cosentino, X. Dumusque, P. Figueira, D. B. Guenther, A. Harutyunyan, C. Hellier, R. Kuschnig, M. Lopez-Morales, M. Mayor, G. Micela, A. F. J. Moffat, M. Pedani, D. F. Phillips, G. Piotto, D. Pollacco, D. Queloz, K. Rice, R. Riddle, J. F. Rowe, S. M. Rucinski, D. Sasselov, D. Ségransan, A. Sozzetti, A. Szentgyorgyi, C. Watson, W. W. Weiss, Characterizing K2 Planet Discoveries: A Super-Earth Transiting The Bright K Dwarf Hip 116454, The Astrophysical Journal 800 (2015) 59. doi:10.1088/0004-637x/800/1/59.

[55] S. Luszcz-Cook, K. D. Kleer, I. D. Pater, M. Adamkovics, H. Hammel, Retrieving Neptune's aerosol properties from Keck OSIRIS observations. I. Dark regions, Icarus 276 (2016) 52–87. doi:10.1016/j.icarus.2016.04.032.

[56] Y. Matviychuk, E. V. Harbou, J. D. Holland, An experimental validation of a Bayesian model for quantification in NMR spectroscopy, Journal of Magnetic Resonance 285 (2017) 86–100. doi:10.1016/j.jmr.2017.10.009.

[57] S. Lunderman, M. Morzfeld, F. Glassmeier, G. Feingold, Estimating parameters of the nonlinear cloud and rain equation from a large-eddy simulation, Physica D: Nonlinear Phenomena 410 (2020) . doi:10.1016/j.physd.2020.132500.

[58] L. Alawieh, J. Goodman, J. B. Bell, Iterative construction of Gaussian process surrogate models for Bayesian inference, Journal of Statistical Planning and Inference 207 (2020) 55–72. doi:10.1016/j.jspi.2019.11.002.

[59] J. Zhang, M. D. Shields, On the quantification and efficient propagation of imprecise probabilities resulting from small datasets, Mechanical Systems and Signal Processing 98 (2018) 465–483. doi:10.1016/j.ymssp.2017.04.042.

[60] N. H. Paulson, B. J. Bocklund, R. A. Otis, Z. K. Liu, M. Stan, A Stochastic Regression Approach to Analyzing Thermodynamic Uncertainty in Chemical Speciation Modeling, Acta Materialia 174 (2019) 9–15. doi:10.1021/es0523035.s001.

[61] F. P. Hollick, V. Gori, C. A. Elwell, Thermal performance of occupied homes: A dynamic grey-box method accounting for solar gains, Energy and Buildings 208 (2020) . doi:10.1016/j.enbuild.2019.109669.

[62] P. Wagner, R. Fahrni, M. Klippel, A. Frangi, B. Sudret, Bayesian calibration and sensitivity analysis of heat transfer models for fire insulation panels, Engineering Structures 205 (2020) . doi:10.1016/j.engstruct.2019.110063.

[63] J. Zhang, S. Termaath, M. D. Shields, Imprecise global sensitivity analysis using bayesian multimodel inference and importance sampling, Mechanical Systems and Signal Processing 148 (2021) . doi:10.1016/j.ymssp.2020.107162.

[64] J. H. Gallier, *Geometric methods and applications: For computer science and engineering*, Springer Science and Business Media, 2012, ISBN: 978-1461301370.

[65] Y. Wang, J. Solomon, Intrinsic and extrinsic operators for shape analysis, Handbook of Numerical Analysis Processing, Analyzing and Learning of Images, Shapes, and Forms: Part 2 (2019) 41–115. doi:10.1016/bs.hna.2019.08.003.

[66] T. Lampart, Implementation and performance comparison of an ensemble sampler with affine invariance, Technical Report, MOSAIC Group, Institute of Theoretical Computer Science, Department of Computer Science, ETH Zurich (2012).

[67] P. H. Garthwaite, Y. Fan, S. A. Sisson, Adaptive optimal scaling of Metropolis–Hastings algorithms using the Robbins–Monro process, Com-

munications in Statistics - Theory and Methods 45 (2016) 5098–5111. doi:10.1080/03610926.2014.936562.

[68] S. T. Thorton, J. B. Marion, *Classical Dynamics of Particles and Systems*, 5th Edition, Brooks/Cole, 2004, ISBN: 978-0534408961.

[69] D. Himmelblau, *Applied Nonlinear Programming*, McGraw-Hill, 1972, ISBN: 0070289212.

[70] C. Safta, M. Khalil, H. N. Najm, Transitional Markov Chain Monte Carlo Sampler in UQTk, SANDIA Report (2020) . doi:10.2172/1606084.

[71] P. Liang, J. E. Mottershead, F. A. DiazDelaO, Model Updating with the Kriging Predictor: Effect of Code Uncertainty, Proceedings of the International Conference on Noise and Vibration Engineering 2016 (ISMA 2016) and International Conference on Uncertainty in Structural Dynamics (USD 2016), Leuven, Belgium 1 (2016) 4363–4375.

[72] H. H. Khodaparast, J. E. Mottershead, K. J. Badcock, Interval model updating with irreducible uncertainty using the Kriging predictor, Mechanical Systems and Signal Processing 25 (2011) 1204–1226. doi:10.1016/j.ymssp.2010.10.009.

[73] J. J. Jenq, W. Li, Feedforward backpropagation artificial neural networks on reconfigurable meshes, Future Generation Computer Systems 14 (1998) 313–319. doi:10.1016/s0167-739x(98)00036-3.

[74] J. He, J. W. Jones, W. D. Graham, M. D. Dukes, Influence of likelihood function choice for estimating crop model parameters using the generalized likelihood uncertainty estimation method, Agricultural Systems 103 (5) (2010) 256–264. doi:10.1016/j.agsy.2010.01.006.

[75] S. K. Zak, K. Beven, B. Reynolds, Uncertainty in the estimation of critical loads: A practical methodology, Water, Air, and Soil Pollution 98 (3-4) (1997) 297–316. doi:10.1007/bf02047040.

[76] K. Beven, A. Binley, The future of distributed models: Model calibration and uncertainty prediction, Hydrological Processes 6 (3) (1992) 279–298. doi:10.1002/hyp.3360060305.

[77] R. Romanowicz, K. J. Beven, J. Tawn, Evaluation of predictive uncertainty in non-linear hydrological models using a bayesian approach, In V. Barnett and K. F. Turkman (Eds.), Statistics for the Environment II, New York, Wiley (1992) 297–317.

[78] R. Romanowicz, K. J. Beven, J. Tawn, Bayesian calibration of flood inundation models, In M. G. Anderson, D. E. Walling and P.D. Bates (Eds.), Flood Plain Processes, Wiley, Chichester (1996) .

[79] X. Wang, X. He, J. R. Williams, R. C. Izaurralde, J. D. Atwood, Sensitivity And Uncertainty Analyses Of Crop Yields And Soil Organic Carbon Simulated With Epic, Transactions of the ASAE 48 (3) (2005) 1041–1054. doi:10.13031/2013.18515.

[80] F. Ruggeri, D. R. Insua, J. Martín, Robust bayesian analysis, Handbook of Statistics Bayesian Thinking - Modeling and Computation (2005) 623–667. doi:10.1016/s0169-7161(05)25021-6.

[81] J. O. Berger, E. Moreno, L. R. P. *et al.*, An overview of robust Bayesian analysis, Test 3 (1994) 5–124. doi:10.1007/BF02562676.

[82] S. Ferson, V. Kreinovick, L. Ginzburg, F. Sentz, Constructing Probability Boxes and Dempster-Shafer Structures, Sandia National Laboratories 4015 (2002) . doi:10.2172/809606.

[83] M. Beer, S. Ferson, V. Kreinovich, Imprecise probabilities in engineering analyses, Mechanical Systems and Signal Processing 37 (2013) 4–29. doi:10.1016/j.ymssp.2013.01.024.