A Gibbs sampling algorithm for structural modal identification under seismic excitation

Binbin Li a, Armen Der Kiureghian b, Siu-Kui Au c

aInstitute for Risk & Uncertainty andCenter for Engineering Dynamics, University of Liverpool,

Liverpool, L69 3GH, UK. Email: bbl@liverpool.ac.uk

bDepartment of Civil and Environmental Engineering, University of California, Berkeley,

CA 94720, USA. E-mail: adk@ce.berkeley.edu

American University of Armenia, Yerevan 0019, Armenia, E-mail: President@aua.am

cInstitute for Risk & Uncertainty andCenter for Engineering Dynamics, University of Liverpool,

Liverpool, L69 3GH, UK. Email: siukuiau@liverpool.ac.uk

**Abstract**: Identification of structural modal parameters under seismic excitation using operational modal analysis (OMA) is a challenging task because it violates the basic assumptions of OMA: linear time-invariant model, stationary white noise input and adequately long data. The consequence is significant uncertainties associated with the identified modal parameters. This study aims at developing an algorithm to quantify these uncertainties from a Bayesian perspective. Representing the structure and the seismic excitation by a state-space model, a probabilistic OMA scheme is formulated. The analytical solution for the posterior statistics is not achievable, and a Gibbs sampling algorithm is developed to provide an efficient and robust numerical solution appropriate for practical applications. The performance of the proposed method is validated by identifying a shear-type building using simulated response data under four recorded earthquake motions, and a supertall building - the One Rincon Hill in San Francisco using field-recorded data under seismic and ambient excitations. The computed posterior distributions of modal parameters represent the knowledge extracted from the measured data; they can be reliably used for model validation and health monitoring.

**Keywords**: Operational modal analysis; Uncertainty quantification; Seismic excitation; State-space model; Gibbs sampling

# Introduction

Structural modal parameters, i.e., natural frequencies, damping ratios and mode shapes, are central to earthquake-resistant design and retrofit of structures [1]. Modal identification aims at identifying these parameters for a structure using measured vibration data. Once modal parameters are identiﬁed, they can be used for model validation and health monitoring. For example, changes in the model may be effected to better match modal predictions, and changes in modal parameters can be used to detect the location and severity of damage [2–4].

When a structure is subjected to a ground excitation, such as in an earthquake event, modal identification can be performed using the measured ground motion as input and the measured structural response as output [5–8]. This approach is known as Experimental Modal Analysis (EMA) [9]. In contrast, Operational Modal Analysis (OMA) aims at modal identification by using only the measured structural response. In general, EMA produces more accurate identification than OMA since more information can be extracted from the measured input-output pair than just the output. However, the motion of the base measured at a limited number of locations (usually just one or two) may not give an adequate description of the input, i.e., the measured response of the structure can also be attributed to effects other than the measured motions at the base; and this leads to modeling errors. Furthermore, in many instances, the seismic excitation is not measured, rendering the EMA inapplicable.

During the past three decades, OMA has drawn a great deal of attention [10–12]. For example, the stochastic subspace identification (SSI) [13,14], the blind modal identification [15–17], the Ibrahim time domain method [18], and Frequency-Domain Decomposition [19,20] have all been either directly applied or modified for modal identification under seismic excitation. Modeling the unmeasured seismic excitation as a stochastic process, OMA offers an alternative approach to extract structural modal parameters without recording the base motion.

Although there are several successful applications, applying OMA to identify structural modal parameters under seismic excitation is still a challenging task. The conventional assumptions behind OMA are that the structural model is linear and time-invariant, that the excitation is stationary and white noise, and that the measured data is adequately long (ideally hundreds of times longer than the natural period of the target mode). Furthermore, the identification is typically done for low-amplitude vibrations, which, due to the amplitude-dependence of the structural damping [21], normally corresponds to small damping ratios, i.e., of the order of 0.5-2%. However, in the case of response to an earthquake, the structure may exhibit nonlinearity and variation in time; the input is nonstationary and has colored characteristics that may not be easily identified, especially when they are close to structural modes; and measured records are relatively short, directly affecting the achievable precision. Due to the violation of the basic assumptions of OMA, the identified modal parameters under seismic excitation may be associated with large uncertainties. Though not all OMA methods assume stationarity of the input, e.g., the Blind modal identification [15–17], the large identification uncertainty may still exist because of structural nonlinearity, time variation and short duration. Therefore, it is necessary to quantify the uncertainty in the identified modal parameters for a reliable application of model validation and health monitoring.

Uncertainty quantification of modal parameters in OMA has been studied from both frequentist and Bayesian perspectives. The frequentist approach assumes the existence of true parameter values and adopts an estimator as a proxy. The identification uncertainty refers to the ensemble variability of the estimator over repeated experiments. Uncertainty quantification in the frequentist approach is mainly based on maximum likelihood estimation (MLE) and perturbation analysis. For example, El-Kafafy et al. [22] calculated uncertainty intervals based on MLE; Reynders et al. [23] applied perturbation analysis to construct the uncertainty bounds on modal parameters obtained from SSI. On the other hand, the Bayesian approach provides a rigorous means for quantifying the uncertainty of parameters by regarding modal identification as an inference problem, where probability is used as a measure for the plausibility of outcomes, given a model of the system and measured data [24]. Yuen and Katafygiotis originally applied the Bayesian approach for OMA both from the time [25,26] and frequency domains [27,28], and Au [29] developed a fast algorithm that allows practical implementation.

This paper aims at developing a Bayesian approach for modal identification of civil structures under small or moderate seismic excitations, where the structural behavior is approximately linear and time-invariant. First, a probabilistic model of OMA is developed, taking advantage of a state-space representation of the equations of motion with the unmeasured base motion included in the model as a stochastic process. Next, Gibbs sampling is applied for efficiently calculating the posterior statistics. Conditional distributions are theoretically derived and a robust sampling implementation is provided. The performance of the proposed method is examined by using synthetically generated response data for a 10-story, shear-type building model subjected to four earthquake records. To illustrate the applicability of the method to full-scale civil structures, seismic response data from the One Rincon Hill in San Francisco are analyzed and the estimated modal parameters are compared with estimates obtained from ambient vibrations.

# Problem formulation

## The physical model

For a discretized, linear, time-invariant dynamical system with degrees of freedom (DoFs), the equation of motion under a base motion is represented as

|  |  |  |
| --- | --- | --- |
|  |  | () |

where , and are the mass, damping and stiffness matrices, respectively; , and are the nodal displacement, velocity and acceleration responses relative to the ground, respectively, with and being the initial relative displacement and velocity vectors; is the ground acceleration vector, and is the corresponding influence matrix relating the nodal DoFs of the structure to the input DoFs. The equation of motion in Eqn. (1) is equivalent to the continuous-time state-space model (SSM)

|  |  |  |
| --- | --- | --- |
|  |  | () |

with

|  |  |  |
| --- | --- | --- |
|  | , , | () |

where and represent -by- zero and identity matrices, respectively. is usually called the model order.

In the above SSM, the state variable includes the relative displacements and velocities at all DoFs, but in practice one usually measures the ‘total’ values, not those relative to the ground. In addition, only a few of these quantities can be directly measured due to limited instrumentation. In practice, nodal accelerations are easy to measure with high resolution. Hence, we only consider acceleration measurements, which are expressed through an observation equation

|  |  |  |
| --- | --- | --- |
|  |  | () |

where , , in which is a selection matrix that defines the DoFs of the structure at which acceleration measurements are made.

In practice, the available data is in a discrete form. Therefore, it is necessary to convert Eqn. (2) into a discrete model, where is taken as constant, i.e., , within the small time interval , to obtain

|  |  |  |
| --- | --- | --- |
|  |  | () |

where we have [10]

|  |  |  |
| --- | --- | --- |
|  | , | () |

Correspondingly, the discretized observation equation becomes

|  |  |  |
| --- | --- | --- |
|  |  | () |

with .

Since our objective in the OMA is to identify modal parameters, it is necessary to connect the modal parameters with the state transition matrix and observation matrix . An eigenvalue decomposition of

|  |  |  |
| --- | --- | --- |
|  |  | () |

where and denote the complex conjugate and conjugate transpose of , respectively, yields the continuous time eigenvalues , modal frequencies and damping ratios [10] according to

|  |  |  |
| --- | --- | --- |
|  |  | () |

The mode shapes of the structure for the measured DoFs are given by

|  |  |  |
| --- | --- | --- |
|  |  | () |

The obtained mode shape in Eqn. (10) is a complex vector. However, if the damping matrix is in classical form, i.e., satisfying , the mode shape can be represented in real form without any approximation because all elements are either in-phase or 180-degrees out of phase. In this paper, we do not require classical damping so that the modal frequencies and damping ratios in Eqn. (9) are only nominal values and mode shapes are generally complex, but we apply a post-processing strategy [30] to obtain approximated real mode shapes in displaying the identification results.

## The probabilistic model

The structural dynamic system is always subject to various kinds of errors: Eqn. (5) cannot exactly predict the structural behavior due to the existence of modeling errors, e.g., the nonlinearity and time-variance of the structure, and Eqn. (7) must account for measurement errors that are inevitably present. Considering these effects, we model the equations of motion by a stochastic SSM as

|  |  |  |
| --- | --- | --- |
|  |  | () |

where and represent the modeling and measurement errors, respectively. Furthermore, the base motions are not measured in the OMA and, therefore, they are modeled as random processes. The base motions can be approximated by time-variant autoregressive moving-average (ARMA) [31] or time-invariant ARMA models with amplitude modulation [32]. Here, we adopt the time-invariant ARMA model for its simplicity. This may introduce additional modeling errors, but it suffices for our purpose, as validated by subsequent empirical studies. Recalling the equivalence between the ARMA model and the SSM [33], we only need to augment our existing SSM in Eqn. (11) to include the base motion model, i.e.,

|  |  |  |
| --- | --- | --- |
|  |  | () |

where in which is the order of the base motion model and is the stochastic term for the ground motion. More compactly, we can write the above model by a stochastic SSM:

|  |  |  |
| --- | --- | --- |
|  |  | () |

with

|  |  |  |
| --- | --- | --- |
|  | ,, , | () |

The effect of introducing the model for the base motion is that we have to use a higher model order in modal identification and we must remove spurious modes resulting from predominant frequencies in the ground motion through post-processing.

The joint distribution of and is assumed to be the multivariate normal with zero-mean and unknown covariance matrix for , i.e., . This assumption can be justified by the principle of maximum entropy [24]. As a consequence, the joint distribution of given , and is also a multivariate normal with the probability density function (PDF)

|  |  |
| --- | --- |
|  | () |

where ‘’ denotes the determinant of the matrix.

Given the PDF in Eqn. (15), we choose a multivariate normal distribution as the prior of the initial response and a matrix normal inverse Wishart distribution [34] as the joint prior distribution of and so that the formulated probabilistic model belongs to the conjugate-exponential family [35]. These prior PDFs are defined by

|  |  |  |
| --- | --- | --- |
|  |  | () |

where represents the multivariate gamma function and denotes the matrix trace; the mean and the covariance matrix are the parameters in the multivariate normal distribution; the degree of freedom and scale matrix are the parameters in the inverse Wishart distribution; and the mean and the right-covariance matrix are parameters in the matrix normal distribution. These quantities are parameters of different PDFs, but hyper-parameters of the overall probabilistic model, so hereafter we call them hyper-parameters.

Since the state variables cannot be directly measured, we call them latent variables. Correspondingly, we call the measured structural responses as observed variables. Collecting all the observed variables , the latent variables , and the unknown parameters **,** as well as their associated PDFs, yields the probabilistic model for the OMA expressed by the joint distribution

|  |  |  |
| --- | --- | --- |
|  |  | () |

where the PDFs in the right hand side are given in Eqns. (15) and (16). Once the structural responses are recorded, it is possible to infer the posterior distribution of unknown parameters . Due to the complexity of the model, exact Bayesian inference is not feasible. As an alternative, in the following section, we apply Gibbs sampling to tackle the problem.

# Gibbs sampling for OMA

Gibbs sampling [36,37] is a popular Markov chain Monte Carlo method. The basic idea is to generate posterior samples by sweeping through each variable (or block of variables) and draw from its conditional distribution, with the remaining variables fixed to their current values. Gibbs sampling is simple, and it is straightforward to verify its theoretical validity [38]. It is particularly well-suited to the proposed probabilistic model of OMA, because the conditional distributions and belong to standard conjugate-exponential family of distributions, so that they can be efficiently sampled.

In Gibbs sampling, given the measurements , we need to derive the conditional distributions and and iteratively sample from them. Thus, using superscript to represent the iteration step, we proceed as follows:

(1) Given and and the measurement , generate a sample of the latent variables according to

|  |  |
| --- | --- |
|  |  |

(2) Given , generate a sample of the unknown parameters according to

|  |  |
| --- | --- |
|  |  |

(3) Calculate the modal parameters , and from based on the procedure introduced in Section 2.1.

In the following two subsections, we develop an algorithm to draw samples from the conditional distributions and . A robust implementation is presented in Subsection 3.3.

## Sampling latent variables

Considering the joint distribution in Eqn. (17), the conditional distribution of the latent variables is

|  |  |
| --- | --- |
|  | () |

This is a multivariate normal distribution, but for the present model directly sampling from it is not feasible because of the extremely demanding task of computing the matrix inverse , when is large. As an alternative, we employ the forward-filtering-backward-sampling (FFBS) algorithm [33] to take advantage of the conditional Markov property [39].

The conditional distribution of the latent variables can be written as

|  |  |  |
| --- | --- | --- |
|  |  | () |

where we have used the conditional Markov property in the last line. This factorization highlights the fact that we can sample from by using a backward sampling strategy: First, draw a sample from ; then, conditioned on , draw from the conditional density and continue in this fashion until . In particular, is the distribution available from the Kalman filter recursions [40], which is shown in Algorithm 1 below. As for the PDF , recall the Markov property

|  |  |
| --- | --- |
|  | () |

and the fact that the conditional joint distribution of and given , and is a multivariate normal distribution with the mean vector and covariance matrix given by [39]

|  |  |  |
| --- | --- | --- |
|  | , | () |

where the quantities and for are calculated in the forward filtering step as shown in Algorithm 1 below. Using the property of the multivariate normal distribution, we know that the conditional distribution of given , , and is still multivariate normal with the mean and covariance

|  |  |  |
| --- | --- | --- |
|  |  | () |

where . Therefore, all the latent variables can be sampled from multivariate normal distributions.

As a summary, all steps of the FFBS algorithm are listed in Algorithm 1.

|  |
| --- |
| 1) Initialization  Define , ,  ,  and set and  2) Forward inference (Kalman filter)  For to  Measurement update        Time update      End For  3) Backward sampling  Sample  For to    Sample  where    End For |
| Algorithm 1: Forward-filtering-backward-sampling. |

## Sampling unknown parameters

Sampling the unknown parameters is a relatively easy task because the conjugate prior is used so that the posterior is again the matrix normal inverse Wishart distribution. From the joint distribution in Eqn. (17), we have

|  |  |
| --- | --- |
|  | () |

where

|  |  |
| --- | --- |
| ,  , | () |

Once the hyper-parameters are known, it is straightforward to sample from the standard distributions. We first sample , then conditioned on , we sample , which is equivalent to sampling from the multivariate normal distribution [34], where ‘’ means stacking the columns of the matrix into a column vector and ‘’ denotes the Kronecker product.

## Robust implementation

Since the matrices , , , and must remain symmetric and positive semi-definite at all iterations, a robust implementation of the sampling strategy is essential. We found that a naïve implementation directly following the steps described in Subsections 3.1 and 3.2 suffered from numerical errors. Here, we apply the square-root filtering strategy [40] to overcome the accumulated numerical error and guarantee semi-definiteness of the covariance matrices.

In the square-root filtering algorithm, we need the square-root of a symmetric and positive semi-definite matrix. Here, we require it to be an upper-triangular matrix, which can be obtained via Cholesky decomposition [41], i.e., for in Algorithm 1

|  |  |
| --- | --- |
|  | () |

For the other square-root matrices in the initialization step of the algorithm, we perform the following Cholesky decomposition

|  |  |
| --- | --- |
|  | () |

One can show that

|  |  |
| --- | --- |
| , , | () |

In the forward simulating step, we apply the QR decomposition [41] to obtain the square-root matrices. First, in the measurement update, we employ the QR decomposition

|  |  |
| --- | --- |
|  | () |

Taking advantage of the unitary nature of , one can verify that

|  |  |
| --- | --- |
| , | () |

Next, for the time update, employing the square-root in the QR decomposition

|  |  |
| --- | --- |
|  | () |

yields

|  |  |
| --- | --- |
| , | () |

For the backward sampling, given can be sampled as

|  |  |
| --- | --- |
|  | () |

where is a vector of standard normal random variables, i.e., zero-mean and identity-covariance. Then, again applying the QR decomposition

|  |  |
| --- | --- |
|  | () |

yields

|  |  |
| --- | --- |
| , | () |

Therefore, we can sample by first sampling an -dimension standard normal random vector and setting

|  |  |
| --- | --- |
|  | () |

For the robust implementation of the sampling of the unknown parameters with hyper-parameters , , and shown in Eqn. (24), we perform the Cholesky decomposition

|  |  |
| --- | --- |
|  | () |

where

|  |  |
| --- | --- |
| , , | () |

Equating the sub-matrices yields

|  |  |
| --- | --- |
| , , | () |

For the purpose of efficiently sampling after sampling based on and , we apply the following procedure [34]: Perform the Cholesky decomposition , then sample a standard normal matrix and set

|  |  |
| --- | --- |
|  | () |

The procedure for robust implementation of the Gibbs sampler is summarized in Algorithm 2.

|  |
| --- |
| 1) Initialization  Set  and  , ,  For to  2) Robust latent variables sampling  For to  Measurement update    ,  Time update    ,  End For  where  For to    where  End For  3) Robust unknown parameters sampling  Compute matrix , , defined in Eqn. (37)    and  where  4) Modal parameters computation following the procedure provided in Section 2.1.  End For |
| Algorithm 2: Robust implementation of forward-filtering-backward-sampling. |

# Empirical studies

This section presents empirical studies of the proposed method via a numerical shear-type building model and a real structure - the One Rincon Hill in San Francisco, California. For the prior means, we choose estimates based on SSI [42]; for the prior variances we assume large values so that the prior distributions can be regarded as non-informative. Taking advantage of parallel computing [43], four independent Markov chains are used in the Gibbs sampling with 1000 samples in each chain. The Gelman-Rubin measure [44], which estimates the potential decrease in the between-chains variance with respect to (w.r.t.) the within-chain variance, is applied to assess the convergence of the generated samples. As suggested by Brooks and Gelman [45], if for all model parameters, one can be fairly confident that convergence has been reached. Otherwise, longer chains or other means for improving the convergence may be needed. To be more reassuring, we apply a more stringent condition in the examples below.

## Example with synthetic data

We first consider the modal identification of an idealized 10-story shear-type building model using synthetically generated acceleration responses under four different seismic excitations. Since the true modal parameters are known a priori, it is possible to use the model to validate the performance of the proposed algorithm.

The properties of the building are listed in Table 1. This model has been considered by Pioldi and Rizzi [20] to compare two OMA algorithms for modal identification under seismic excitations, but we have modeled damping by directly specifying the damping ratios rather than assuming Rayleigh damping. In order to compare the performance of our algorithms under different seismic excitations, four different earthquake records are selected, as shown in Table 2 and Figure 1. These records are downloaded from the Center of Engineering Strong Motion Data (CESMD) online database [46]. They have been selected as representative of a wide variety of seismic events with rather different characteristics, but all with relatively short durations. Synthetic acceleration responses are calculated using the discretized SSM in Eqn. (5), and then contaminated by a Gaussian white-noise process with two-sided root power spectral density (PSD) of to model the measurement noise. Since the highest frequency of the model is *Hz*, the raw data are down sampled from 50 *Hz* to *Hz*. In this example, acceleration responses at all stories are used for modal identification, and their root singular value (SV) spectra are shown in Figure 2. The root SV spectrum plots the square root of eigenvalues of the PSD matrix, giving a rough idea of natural frequencies and the quality of data. More specifically, the peaks in the top line indicate natural frequencies, and the ratio of the top to second top line at the natural frequency is approximately the square root of the modal signal-to-noise ratio [12].

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table : Model properties of the 10-story shear-type building | | | | | | | | | | |
| Story | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| Stiffness [×103 *kN*/*m*] | 62.47 | 59.26 | 56.14 | 53.02 | 49.91 | 46.79 | 43.70 | 40.55 | 37.43 | 34.31 |
| Mass [×103 *kg*] | 179 | 170 | 161 | 152 | 143 | 134 | 125 | 116 | 107 | 98 |
| Mode | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| Frequency [*Hz*] | 0.50 | 1.33 | 2.15 | 2.93 | 3.65 | 4.29 | 4.84 | 5.27 | 5.59 | 5.79 |
| Damping ratio [*%*] | 5 | 4 | 3 | 2 | 1 | 1 | 1 | 1 | 1 | 0.8 |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table : Features of selected earthquake records | | | | | | | | |
| Earthquake | Date | Magnitude  [*Mw*] | Station | Hypocentral  distance[*km*] | PGA[*g*] | Component | Duration  [*sec*] | Sampling  frequency[*Hz*] |
| El Centro | 05/18/1940 | 6.9 | El Centro | 16.9 | 0.349 | 0 | 54 | 50 |
| Tabas | 09/16/1978 | 7.3 | Tabas | N.A. | 0.932 | 344 | 43 | 50 |
| Loma Prieta | 10/18/1989 | 7.0 | Capitola | 20.1 | 0.399 | 0 | 40 | 50 |
| Northridge | 01/17/1994 | 6.7 | Northridge | 19.1 | 0.453 | 180 | 60 | 50 |

We apply Gibbs sampling to the synthetic data to extract the modal parameters with the initial setting given at the beginning of this section. Figure 3 illustrates a typical convergence process of the generated samples. It takes 600 steps for the Gelman-Rubin measure to be less than . Since only the last half of the samples is used in computing , it follows that 300 samples are required for convergence. Therefore, the first 300 samples in each chain are discarded, giving rise to 2800 posterior samples in total. To incorporate the effect of the ground motion, the order of the state transition matrix is set to , which is much higher than the true order of the structural model (). As a consequence, we need to remove spurious seismic modes in the post-processing. It turns out that it is not a serious issue in this example because only the structural modes appear consistently in all samples, as evidenced in Figure 4, where the left figure shows the stabilization diagram [10] w.r.t. generated samples, and the right figure plots the histogram of identified frequencies and damping ratios. Though there are many spurious modes, only those structural modes cluster together and consistently show in all samples. The whole sampling process takes 35.5 min on a Digital Storm laptop with Intel® Core™ i7 CPU @2.50 GHz and RAM 16.0 GB. By analyzing the Gibbs sampling algorithm, the computation time is proportional to the data length , the required number of samples and the cube of the system order . Therefore, the controlling factor is the system order, which should be set not too large for computational efficiency.

|  |
| --- |
|  |
| Figure 1: Time history and root PSD of adopted earthquake records. |

|  |
| --- |
|  |
| Figure 2: Root SV (singular value) spectrum of synthetically generated responses. |

|  |
| --- |
|  |
| Figure 3: Convergence process of the Gibbs sampling: El Centro earthquake. |

|  |  |
| --- | --- |
|  |  |
| a) Stabilization diagram | b) histogram |
| Figure 4: Removing spurious modes: El Centro earthquake. | |

The Posterior samples of natural frequencies and damping ratios identified when using the El Centro record are shown in Figure 5. As can be seen, the true parameter values are close to the cloud of sampled values, especially the frequencies. Consistent with common findings, the natural frequencies are identified with high precision, but estimated damping ratios have much larger variability; longer data are required to improve the precision. Since the data set is short and the model assumption on the seismic excitation is violated, the posterior samples need not necessarily cover the true values. That is the case, for example, with the fourth mode. Nevertheless, the results are sufficiently good to give representative values of the modal parameters. More importantly, the full posterior distributions are obtained and can be reliably used for further applications.

|  |  |
| --- | --- |
|  |  |
|  |  |
| Figure 5: Identified frequencies and damping ratios: first four modes, El Centro earthquake. | |

|  |
| --- |
|  |
| Figure 6: Identified frequencies under four seismic excitations  Normalized f = identified frequency/true frequency. |

|  |
| --- |
|  |
| Figure 7: Identified damping ratios under four seismic excitations  Normalized ζ = identified damping ratio/true damping ratio. |

|  |
| --- |
|  |
| Figure 8 MAC between the true and identified mode shapes  subjected to four seismic excitations. |

The natural frequencies, damping ratios and mode shapes identified with the four seismic records are illustrated via boxplots in Figure 6-Figure 8. In these plots, the natural frequencies and damping ratios are normalized w.r.t. their true values, and the mode shapes are represented in terms of the modal assurance criterion (MAC) with respect to the true mode shapes [11]. Most identified ranges of natural frequencies and damping ratios cover the true values, and most MACs are great than 0.99, thus validating the performance of the proposed algorithm. The uncertainty in the estimated natural frequency decreases with increasing modal frequency, which is explained by the increase of the effective data length (= data duration/natural period), i.e., the effective cycles of vibration. This suggests that, given a specified precision, the fundamental mode is more challenging to identify among all modes that are well excited [47]. Since some modes are not adequately excited and due to the existence of modeling error, the posterior distribution may not cover the true value and may exhibit a large uncertainty. For example, under the Loma Prieta record, the estimated mean damping ratio is 2.5%, which is far from the true value, and the corresponding MAC is only 0.44. In general, since damping ratios are amplitude dependent and their estimation is associated with large errors, one can only judge whether the magnitudes of estimated values are reasonable. Under these conditions, the results should be interpreted with these aspects in mind. Besides the modal parameters presented above, the system state and modeling/measurement variance are also identified in terms of samples. There is potential use of them, e.g., for damage detection, but we do not illustrate them here because our main interest is in modal parameters.

## Example with field data

The second example is the One Rincon Hill, a 64-story reinforced-concrete, shear-wall building in San Francisco, California. A 72-channel seismic monitoring system (Figure 9) was installed to stream real-time acceleration data throughout the building. Both seismic (South Napa earthquake, 14/08/2014, *Mw* = 6.0) and ambient vibration data (27/12/2012) are available in the CESMD online database with station number 58389 [46]. This building is used to investigate the practical applicability of the proposed algorithm and to compare its performance under seismic and ambient excitations.

|  |
| --- |
|  |
| Figure 9: Seismic monitoring system of ORTH and sensors adopted in this research [46]. |

In order to capture the whole in-plane motion of the floor, we only adopt levels with 3 CSMIP sensors, one in north-south (N/S) and two in east-west (E/W) directions of each floor, resulting in 18 sensors in total, as shown in Figure 9. The seismic response data have a sampling frequency of 100 *Hz* and a duration of 90 *sec*, while the ambient vibration data are sampled at 200 *Hz* for 230 *sec*. Two typical sensor records as well as the root SV spectra of the overall dynamic responses are illustrated in Figure 10. Evidently, large differences exist between the estimates based on the seismic and ambient vibration excitations: The seismic data exhibits high-amplitude and nonstationarity, and its frequency content is not as rich as ambient data. In particular, structural modes at about 0.7, 2.0 and 3.7 *Hz* are not well excited by the South Napa earthquake.

|  |  |
| --- | --- |
|  |  |
| a) South Napa earthquake | b) Ambient |
| Figure 10: Time history (top) and root SV spectra (bottom). | |

For the sake of comparability and reduction of computational burden, both sets of raw data are resampled down to 25 *Hz*. According to Ref. [30], modes with frequencies below 5 *Hz* are of interest in this study. Applying the Gibbs sampling algorithm, the convergence process of the seismic responses is illustrated in Figure 11, from which 2400 stationary samples (= 4 chains × 600 samples/chain) are obtained based on the Gelman-Rubin measure . The order of the state transition model is set to 60 to account for both the structural, seismic mode and possible mathematical modes. In order to remove spurious modes, the stabilization diagram and the histogram of samples are presented in Figure 12. In this case, it is not easy to detect spurious modes. For example, the spurious mode at 2.78 *Hz* appears to be stable, but the physical modes at 0.7, 2.0 and 3.7 *Hz* appear unstable because these 3 torsional modes are not well excited. Taking advantage of the identification results based on the ambient vibration data, we are able to detect the spurious modes by checking the consistency of each mode in natural frequency and MAC. However, it is not always straight forward to detect the spurious modes, especially when using seismic excitations. The existence of multiple seismic records would greatly facilitate the detection of spurious modes, because physical modes tend to be consistent from earthquake to earthquake (assuming they are excited), but spurious modes may not.

|  |
| --- |
|  |
| Figure 11: Convergence process of the Gibbs sampling: South Napa earthquake. |

|  |  |
| --- | --- |
|  |  |
| a) Stabilization diagram | b) histogram |
| Figure 12: Removing spurious modes: South Napa earthquake. | |

After removing the spurious modes, the identified results are listed in Table 3 and the mean mode shapes of lateral modes are shown in Figure 13. The mean frequencies identified from South Napa earthquake are all smaller than those identified from ambient excitation, while most of the mean damping ratios are bigger. These might be due to the large amplitude of seismic excitation, but we cannot make a conclusive statement because modal parameters can show large variabilities due to environmental effects and the identified damping ratios are associated with large uncertainties. Since the first and second torsional modes are not well excited during the South Napa earthquake, their mode shapes show small coherence within the two data sets, but all other mode shapes are close to each other. The ambient vibration data are much longer than the seismic excitation, and correspondingly the identification uncertainties in natural frequencies and damping ratios are smaller. Again, we see that uncertainties decrease with increasing modal frequency because of the increase in the effective data length.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Table : Identified modal parameters: One Rincon Hill | | | | | | |
| Mode | Natural frequencies [*Hz*] | | Damping ratios [*%*] | | Mode shapes | |
| South Napa | Ambient | South Napa | Ambient | MAC [*%*] | Description |
| 1 | 0.26(1.2) | 0.27(0.67) | 2.18(52) | 1.61(40) | 99.86 | 1st E/W |
| 2 | 0.30(1.2) | 0.30(0.75) | 2.31(48) | 2.20(35) | 98.09 | 1st N/S |
| 3 | 0.68(0.94) | 0.70(0.40) | 2.04(37) | 1.39(30) | 54.60 | 1st torsion |
| 4 | 1.11(0.53) | 1.14(0.49) | 1.50(35) | 2.91(23) | 99.54 | 2nd E/W |
| 5 | 1.26(0.41) | 1.30(0.36) | 1.04(38) | 1.78(21) | 97.45 | 2nd N/S |
| 6 | 1.98(0.47) | 2.04(0.21) | 2.02(24) | 1.11(19) | 14.40 | 2nd torsion |
| 7 | 2.58(0.39) | 2.65(0.22) | 1.13(31) | 1.09(28) | 98.10 | 3rd E/W |
| 8 | 2.79(0.21) | 2.86(0.12) | 0.57(35) | 0.52(24) | 99.25 | 3rd N/S |
| 9 | 3.68(0.29) | 3.73(0.10) | 1.02(26) | 0.46(22) | 97.49 | 3rd torsion |
| 10 | 4.03(0.33) | 4.13(0.17) | 1.24(27) | 1.00(17) | 99.79 | 4th E/W |
| 11 | 4.22(0.49) | 4.34(0.18) | 2.06(24) | 1.04(15) | 97.90 | 4th N/S |
| Note: Mean values with c.o.v. (coefficient of variation) in parenthesis (units: *%*) are used for frequencies and damping ratios; the MAC (modal assurance criterion) is calculated based on the mean of identified mode shapes. | | | | | | |

|  |  |
| --- | --- |
|  |  |
|  |  |
| Figure 13: Mean of first eight lateral mode shapes: South Napa earthquake. | |

|  |  |
| --- | --- |
|  |  |
|  |  |
| a) 1st and 2nd modes | |
|  |  |
|  |  |
| b) 10th and 11th modes | |
| Figure 14: Identified frequencies and damping ratios: South Napa earthquake. | |

An interesting phenomenon of this building is the existence of closely-spaced modes due to similar masses and stiffnesses in the N/S and E/W directions. Four pairs of modes appear within the frequency band of interest. Being different from the well separated modes, the identified modal parameters can be correlated for closely-spaced mode pairs. Scatter plots of two pairs of closely-spaced modes are provided in Figure 14, in which the correlation coefficient is also shown. It can be seen that in some cases the correlation coefficient is not negligible, e.g., the correlation coefficient between damping ratios and is 0.12.

# Conclusions

The structural modal identification problem under seismic excitation is investigated from the perspective of operational modal analysis (OMA). Modeling both the structural and seismic excitation in terms of a state-space model, a probabilistic model is first constructed and then resolved by Gibbs sampling. An efficient and robust implementation of the Gibbs sampling is developed, allowing its use in practical applications. The approach allows not only a point estimate of modal parameters, but also an approximation of the whole posterior distribution by the generated samples. Comparing with previously proposed Bayesian methods [25–29], the present method is a good complement because it does not require the assumption of classical damping. Through careful examination of the proposed algorithm for examples with synthetic and field data, the following conclusions are made:

(1) Though the underlying assumptions of OMA are violated, identification of modal parameters by OMA under seismic excitation is still feasible on the condition of small or moderate intensity excitation, where the structural behavior remains approximately linear and time-invariant;

(2) The identified modal parameters, especially the damping ratios, are associated with large uncertainties due to the short duration of the earthquake record. One should exercise caution when using the estimated modal parameters to detect damage or update a finite element model. Full consideration of the posterior distributions is necessary to properly account for the prediction uncertainty.

(3) The uncertainty in modal parameters decreases with increasing modal frequency because of the increase in the effective data length (=data duration/natural period), i.e., the effective cycles of vibrations. According to the posterior distribution, the estimated modal parameters are almost uncorrelated for well separated modes, but correlation may exist between the estimates for closely-spaced modes.

(4) The detection of spurious modes by use of the stabilization diagram or histogram may still be an issue, especially when the modes are not well excited under seismic excitations. The existence of multiple seismic or ambient vibration records may provide information to remove spurious modes, because only the physical modes are consistently identified. It should be mentioned that the existence of spurious modes has a minimal effect on the identification uncertainties of the structural modes. This has been validated by a parameter analysis, but it is not shown here due to space limitation.

# Acknowledgements

The first two authors acknowledge support from the US National Science Foundation under Grant No. CMMI-1130061. In addition, the first and third authors thank the UK Engineering and Physical Sciences Research Council (Grant EP/N017897/1) for the partial funding support.

# References

1. Moehle JP. *Seismic design of reinforced concrete buildings*. New York: McGraw-Hill Education; 2014.

2. Brownjohn JMW. Structural health monitoring of civil infrastructure. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* 2007; **365**(1851): 589–622. DOI: 10.1098/rsta.2006.1925.

3. Çatbaş FN, Kijewski-Correa T, Aktan AE, editors. *Structural Identification of Constructed Systems: Approaches, Methods, and Technologies for Effective Practice of St-Id*. Reston, VA: American Society of Civil Engineers; 2013. DOI: 10.1061/9780784411971.

4. Goulet JA, Michel C, Kiureghian A Der. Data-driven post-earthquake rapid structural safety assessment. *Earthquake Engineering and Structural Dynamics* 2015; **44**(4): 549–562. DOI: 10.1002/eqe.2541.

5. Luş H, Betti R, Longman RW. Identification of linear structural systems using earthquake-induced vibration data. *Earthquake Engineering & Structural Dynamics* 1999; **28**(11): 1449–1467. DOI: 10.1002/(SICI)1096-9845(199911)28:11<1449::AID-EQE881>3.0.CO;2-5.

6. Kim J, Lynch JP. Subspace system identification of support-excited structures-part I: theory and black-box system identification. *Earthquake Engineering & Structural Dynamics* 2012; **41**(15): 2235–2251. DOI: 10.1002/eqe.2184.

7. Hong AL, Betti R, Lin CC. Identification of dynamic models of a building structure using multiple earthquake records. *Structural Control and Health Monitoring* 2009; **16**(2): 178–199. DOI: 10.1002/stc.289.

8. Ulusoy HS, Feng MQ, Fanning PJ. System identification of a building from multiple seismic records. *Earthquake Engineering & Structural Dynamics* 2011; **40**(6): 661–674. DOI: 10.1002/eqe.1053.

9. Ewins DJ. *Modal testing : theory, practice, and application*. Baldock: Research Studies Press; 2000.

10. Reynders E. System Identification Methods for (Operational) Modal Analysis: Review and Comparison. *Archives of Computational Methods in Engineering* 2012; **19**(1): 51–124. DOI: 10.1007/s11831-012-9069-x.

11. Brincker R, Ventura C. *Introduction to operational modal analysis*. Chichester: Wiley; 2015.

12. Au SK. *Operational modal analysis : modeling, bayesian inference, uncertainty laws*. Singapore: Springer; 2017.

13. Huang CS, Lin HL. Modal identification of structures from ambient vibration, free vibration, and seismic response data via a subspace approach. *Earthquake Engineering & Structural Dynamics* 2001; **30**(12): 1857–1878. DOI: 10.1002/eqe.98.

14. Loh CH, Chao SH, Weng JH, Wu TH. Application of subspace identification technique to long-term seismic response monitoring of structures. *Earthquake Engineering & Structural Dynamics* 2015; **44**(3): 385–402. DOI: 10.1002/eqe.2475.

15. Sadhu A, Hazra B, Narasimhan S. Blind identification of earthquake-excited structures. *Smart Materials and Structures* 2012; **21**(4): 45019. DOI: 10.1088/0964-1726/21/4/045019.

16. Yang Y, Nagarajaiah S. Blind modal identification of output-only structures in time-domain based on complexity pursuit. *Earthquake Engineering & Structural Dynamics* 2013; **42**(13): 1885–1905. DOI: 10.1002/eqe.2302.

17. Ghahari SF, Abazarsa F, Ghannad MA, Taciroglu E. Response-only modal identification of structures using strong motion data. *Earthquake Engineering & Structural Dynamics* 2013; **42**(8): 1221–1242. DOI: 10.1002/eqe.2268.

18. Lin CC, Hong LL, Ueng JM, Wu KC, Wang CE. Parametric identification of asymmetric buildings from earthquake response records. *Smart Materials and Structures* 2005; **14**(4): 850–861. DOI: 10.1088/0964-1726/14/4/045.

19. Pioldi F, Ferrari R, Rizzi E. Earthquake structural modal estimates of multi-storey frames by a refined Frequency Domain Decomposition algorithm. *JVC/Journal of Vibration and Control* 2017; **23**(13): 2037–2063. DOI: 10.1177/1077546315608557.

20. Pioldi F, Rizzi E. Earthquake-induced structural response output-only identification by two different Operational Modal Analysis techniques. *Earthquake Engineering & Structural Dynamics* 2018; **47**(1): 257–264. DOI: 10.1002/eqe.2947.

21. Jeary AP. Damping in tall buildings—a mechanism and a predictor. *Earthquake Engineering & Structural Dynamics* 1986; **14**(5): 733–750. DOI: 10.1002/eqe.4290140505.

22. El-Kafafy M, De Troyer T, Guillaume P. Fast maximum-likelihood identification of modal parameters with uncertainty intervals: A modal model formulation with enhanced residual term. *Mechanical Systems and Signal Processing* 2014; **48**(1–2): 49–66. DOI: 10.1016/j.ymssp.2014.02.011.

23. Reynders E, Pintelon R, De Roeck G. Uncertainty bounds on modal parameters obtained from stochastic subspace identification. *Mechanical Systems and Signal Processing* 2008; **22**(4): 948–969. DOI: 10.1016/j.ymssp.2007.10.009.

24. Beck JL. Bayesian system identification based on probability logic. *Structural Control and Health Monitoring* 2010; **17**(7): 825–847. DOI: 10.1002/stc.424.

25. Yuen KV, Katafygiotis LS. Bayesian time–domain approach for modal updating using ambient data. *Probabilistic Engineering Mechanics* 2001; **16**(3): 219–231. DOI: 10.1016/S0266-8920(01)00004-2.

26. Yuen KV, Beck JL, Katafygiotis LS. Probabilistic approach for modal identification using non-stationary noisy response measurements only. *Earthquake Engineering & Structural Dynamics* 2002; **31**(4): 1007–1023. DOI: 10.1002/eqe.135.

27. Yuen K, Katafygiotis L. Bayesian fast Fourier transform approach for modal updating using ambient data. *Advances in Structural Engineering* 2003; **6**(2): 81–95. DOI: 10.1260/136943303769013183.

28. Katafygiotis LS, Yuen KV. Bayesian spectral density approach for modal updating using ambient data. *Earthquake Engineering and Structural Dynamics* 2001; **30**(8): 1103–1123. DOI: 10.1002/eqe.53.

29. Au SK, Zhang FL, Ni YC. Bayesian operational modal analysis: Theory, computation, practice. *Computers & Structures* 2013; **126**: 3–14. DOI: 10.1016/j.compstruc.2012.12.015.

30. Li B, Der Kiureghian A. Operational modal identification using variational Bayes. *Mechanical Systems and Signal Processing* 2017; **88**: 377–398. DOI: 10.1016/j.ymssp.2016.11.007.

31. Chang MK, Kwiatkowski JW, Nau RF, Oliver RM, Pister KS. Arma models for earthquake ground motions. *Earthquake Engineering & Structural Dynamics* 1982; **10**(5): 651–662. DOI: 10.1002/eqe.4290100503.

32. Ólafsson S, Sigbjörnsson R. Application of arma models to estimate earthquake ground motion and structural response. *Earthquake Engineering & Structural Dynamics* 1995; **24**(7): 951–966. DOI: 10.1002/eqe.4290240703.

33. Shumway RH, Stoffer DS. *Time series analysis and its applications : with R examples*. 3rd ed. New York: Springer Science + Business Media; 2011.

34. Gupta AK, Nagar DK. *Matrix Variate Distributions*. Boca Raton: Chapman & Hall/CRC; 1999. DOI: 10.1198/jasa.2003.s280.

35. Keener RW. *Theoretical statistics : topics for a core course*. New York: Springer; 2010.

36. Geman S, Geman D. Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 1984; **PAMI**-**6**(6): 721–741. DOI: 10.1109/TPAMI.1984.4767596.

37. Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB. *Bayesian data analysis*. Third Edit. Boca Raton: Chapman & Hall/CRC; 2014.

38. Andrieu C, de Freitas N, Doucet A, Jordan MI. An Introduction to MCMC for Machine Learning. *Machine Learning* 2003; **50**(1/2): 5–43. DOI: 10.1023/A:1020281327116.

39. Jordan MI. *An Introduction to Probabilistic Graphical Models*. Lecture Notes: University of California, Berkeley; 2003.

40. Gibson S, Ninness B. Robust maximum-likelihood estimation of multivariable dynamic systems. *Automatica* 2005; **41**(10): 1667–1682. DOI: 10.1016/j.automatica.2005.05.008.

41. Meyer C. *Matrix analysis and applied linear algebra*. Philadelphia: SIAM; 2000.

42. Overschee P, Moor B. *Subspace Identification for Linear Systems : Theory - Implementation - Applications*. Norwell: Kluwer Academic Publishers; 1996.

43. *MATLAB and Parallel Computing Toolbox Release 2016b*. Natick, Massachusetts, United States: The MathWorks, Inc.; 2016.

44. Gelman A, Rubin DB. Inference from Iterative Simulation Using Multiple Sequences. *Statistical Science* 1992; **7**(4): 457–472. DOI: 10.1214/ss/1177011136.

45. Brooks SPB, Gelman AG. General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics* 1998; **7**(4): 434–455. DOI: 10.2307/1390675.

46. CESMD Database. *Center for Engineering Strong Motion Data (CESMD), a Cooperative Database Effort from US Geological Survey (USGS), California Geological Survey (CGS) and Advanced National Seismic System (ANSS)* 2018. http://www.strongmotioncenter.org/ [accessed January 1, 2018].

47. Au SK. Uncertainty law in ambient modal identification - Part I: Theory. *Mechanical Systems and Signal Processing* 2014; **48**(1–2): 15–33. DOI: 10.1016/j.ymssp.2013.07.016.