The Seventh Asian-Pacific Symposium on Structural Reliability and Its Applications (APSSRA2020)
October 4–7 2020, Tokyo, Japan
T. Takada, I. Yoshida & T. Itoi (editors)

Bayesian Data Driven Model for Uncertain Parameters in Structural Health Monitoring

Y.C. Zhu[1] and S.K. Au[2]

[1]*Dynamics Research Group, Department of Mechanical Engineering, The University of Sheffield, UK. Email: yichen.zhu@sheffield.ac.uk*
[2]*School of Civil & Environmental Engineering, Nanyang Technological University, Singapore. Email: ivanau@ntu.edu.sg*

Abstract: In structural health monitoring (SHM), the relationship between dynamic properties of structures and environmental/operational conditions is frequently explored to understand the source of variability, quantify variability under nominal conditions, etc. Such relationship can be complicated and not amenable to analytical modelling. One promising approach is via 'data-driven models'. One typical assumption is that the training data is precisely known, although this need not be the case in SHM when the training data involves dynamic properties that are inevitably identified/derived from vibration measurements, therefore carrying imprecision or identification uncertainty. When the structural response data is measured under working conditions that cannot be directly controlled, the resulting identification uncertainty can be significant and may also vary among different sets of measurements. To address these uncertainties associated with training data, a Bayesian data driven model has been proposed where the identification uncertainty is incorporated via Bayes' rules through the posterior distributions of parameters of interest given the measured response data. This paper focuses on applying the proposed method to SHM of a tall building under a typhoon event, illustrating its feasibility in real applications. A Gaussian process model is used for inferring the relationship between the dynamic properties of the structure identified using operational modal analysis and intensity of wind excitation.
Keywords: Bayesian data driven model, BAYOMA, Structural health monitoring, Gaussian process, Posterior Uncertainty.

1. Introduction

Structural health monitoring (SHM) focuses on assessing the physical conditions of structures based on measured structural response data. It has become an indispensable tool for damage detection and maintenance management, etc. Thanks to the state-of-art SHM systems built into modern structures such as tall buildings and long-span bridges (Chang et al. 2003; Sohn et al. 2003), the serviceability and reliability of structures can be tracked throughout their whole life-cycle. Among others, operational modal analysis (OMA) has become a popular SHM technique where the dynamic properties of structure are identified from vibration response data (Au 2017; Wenzel and Pichler 2005). The health conditions of structures can then be assessed by investigating the relationship dynamic properties of structures and environmental/ operational conditions.

One common way of understanding dynamic behavior of structures under different conditions is to apply 'data-driven' models, which aim at expressing the relationship between the identified dynamic properties and environmental or operational variables based on some training data. Different models have been used in the literature, including Polynomial Chaos Expansions (Spiridonakos et al. 2016); Functionally Pooled model (Kopsaftopoulos et al. 2018) and Kernel Principal Component Analysis (Reynders et al. 2014). Among others, Gaussian Process (GP) has been found to offer an effective means for constructing data driven models (Rasmussen and Williams 2005).

A typical assumption of conventional data driven model is that the training data are known as precious values without uncertainty. This is usually not the case in SHM. Both the dynamic properties of structures and environmental/operational conditions may not be directly measured but are rather identified from measured SHM data, which inevitably carry imprecision. The associated uncertainty depends on the test configurations such as sensor noise and measurement duration, which can also vary among the identified training data points.

To address the forgoing concern, a Bayesian framework has been proposed by the authors (Zhu and Au 2020) which encapsulates the identification uncertainty of training data through its posterior distribution given the measured data when training the data-driven model. Theoretical issues have been investigated in detail and the resulting formula is intuitive and conducive to analysis and computation. In this paper, the framework is applied to OMA data with GP adopted as data driven models. The resulting algorithm is applied to SHM of a tall building under a typhoon event. The dynamic properties exhibiting amplitude dependence in natural frequency and damping ratio under the typhoon event is investigated.

2. Bayesian Data Driven Model for Uncertain Parameters

Conventional data driven models assume that the training data are known precisely. Given the input and output training data $\{x_i, y_i\}$ ($i=1,...,n_s$ where n_s is the number of training points), the hyper parameter $\boldsymbol{\psi}$ associated with the data driven model describing the functional behaviour between input and output are inferred directly. This is usually not the case in SHM, however. Both the input and output training data may not be directly measured but are rather identified based on observations or response measurements from the system; and different training points may have different precision arising from different identification uncertainties. Fig. 1 shows the schematic diagram of the problem described above. The input and output training data $\{x_i, y_i\}$ are both identified from system measurements D_i. Acknowledging limited data and imperfect model, the 'exact' value of the quantity used as training data is unknown. Only the posterior distribution of the input and output data given the measurement $\mathbf{D}$ and in the context of identification model, i.e., $p(\mathbf{X},\mathbf{Y}|\mathbf{D})$ is available. Inferring the data-driven model now requires maximizing the posterior distribution of hyperparameters $\boldsymbol{\psi}$ given the measured data $\mathbf{D}$, i.e., $p(\boldsymbol{\psi}|\mathbf{D})$. A Bayesian data driven model for uncertain training data proposed by the authors (Zhu and Au 2020) has been used in this paper to infer $p(\boldsymbol{\psi}|\mathbf{D})$, which is briefly reviewed in this section.

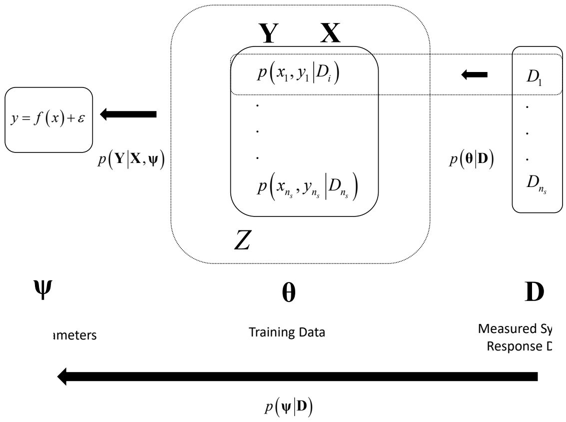

Figure 1. Schematic diagram of proposed framework

Let $\boldsymbol{\theta}$ be a set of parameters identified from the measured data set $\mathbf{D}$, which contains three groups, i.e., $\boldsymbol{\theta}=[\mathbf{X},\mathbf{Y},\mathbf{Z}]$. Here, $\mathbf{X}$ and $\mathbf{Y}$ are used as input and output training data for inferring the data driven model, respectively; $\mathbf{Z}$ contains the remaining parameters identified from $\mathbf{D}$ but not related to the data driven model. Using the theorem of total probability, the marginal distribution of $\boldsymbol{\psi}$ given $\mathbf{D}$ can be expressed as

$$p(\boldsymbol{\psi}|\mathbf{D}) = \int p(\boldsymbol{\psi},\boldsymbol{\theta}|\mathbf{D})d\boldsymbol{\theta} \tag{1}$$

Using Bayes' theorem,

$$p(\boldsymbol{\psi},\boldsymbol{\theta}|\mathbf{D}) = \frac{p(\mathbf{D}|\boldsymbol{\psi},\boldsymbol{\theta})\,p(\boldsymbol{\psi},\boldsymbol{\theta})}{p(\mathbf{D})} \tag{2}$$

Substituting Eq. 2 into Eq. 1 gives

$$p(\boldsymbol{\psi}|\mathbf{D}) = p(\mathbf{D})^{-1}\int p(\mathbf{D}|\boldsymbol{\psi},\boldsymbol{\theta})\,p(\boldsymbol{\psi},\boldsymbol{\theta})d\boldsymbol{\theta} \tag{3}$$

Given $\boldsymbol{\theta}$, the probability distribution of $\mathbf{D}$ can be fully determined via $p(\mathbf{D}|\boldsymbol{\theta})$. The additional information from $\boldsymbol{\psi}$ is therefore redundant, i.e.,

$$p(\mathbf{D}|\boldsymbol{\psi},\boldsymbol{\theta}) = p(\mathbf{D}|\boldsymbol{\theta}) \tag{4}$$

Further applying Bayes' theorem gives,

$$p(\mathbf{D}|\boldsymbol{\theta}) = \frac{p(\boldsymbol{\theta}|\mathbf{D})\,p(\mathbf{D})}{p(\boldsymbol{\theta})} \tag{5}$$

Substituting Eq. 4 and Eq. 5 into Eq. 3 gives

$$p(\boldsymbol{\psi}|\mathbf{D}) = \int p(\boldsymbol{\theta}|\mathbf{D})\,p(\boldsymbol{\psi},\boldsymbol{\theta})\,p(\boldsymbol{\theta})^{-1}d\boldsymbol{\theta} \tag{6}$$

The marginal distribution $p(\boldsymbol{\psi}|\mathbf{D})$ is now expressed in terms of the posterior distribution of $\boldsymbol{\theta}$ given $\mathbf{D}$, i.e. $p(\boldsymbol{\theta}|\mathbf{D})$, which encapsulates the posterior uncertainty of $\boldsymbol{\theta}$. However, the equation still contains information about $\mathbf{Z}$ (inside $\boldsymbol{\theta}$), which is redundant when making inference about $\boldsymbol{\psi}$. It is also necessary to rewrite $p(\boldsymbol{\psi},\boldsymbol{\theta})$ in a more tractable form.

Recall $\boldsymbol{\theta}=[\mathbf{X},\mathbf{Y},\mathbf{Z}]$, $p(\boldsymbol{\psi},\boldsymbol{\theta})$ can be rewritten as

$$p(\boldsymbol{\psi},\boldsymbol{\theta}) = p(\mathbf{X},\mathbf{Y},\mathbf{Z},\boldsymbol{\psi}) = p(\mathbf{X},\mathbf{Y},\mathbf{Z}|\boldsymbol{\psi})\,p(\boldsymbol{\psi}) \tag{7}$$

and note that $\mathbf{Z}$ is not related to $\boldsymbol{\psi}$, it is reasonable to assume that $\{\mathbf{X},\mathbf{Y}\}$ and $\mathbf{Z}$ are conditionally independent for a given $\boldsymbol{\psi}$, i.e.,

$$p(\mathbf{X},\mathbf{Y},\mathbf{Z}|\boldsymbol{\psi}) = p(\mathbf{X},\mathbf{Y}|\boldsymbol{\psi})\,p(\mathbf{Z}) \tag{8}$$

Substituting Eq. 8 into Eq. 7 gives

$$p(\boldsymbol{\psi},\boldsymbol{\theta}) = p(\mathbf{X},\mathbf{Y}|\boldsymbol{\psi})\,p(\boldsymbol{\psi})\,p(\mathbf{Z}) \tag{9}$$

It can also be shown that $\{\mathbf{X},\mathbf{Y}\}$ and $\mathbf{Z}$ are unconditionally independent:

$$\begin{aligned} p(\mathbf{X},\mathbf{Y},\mathbf{Z}) &= \int p(\mathbf{X},\mathbf{Y},\mathbf{Z}|\boldsymbol{\psi})\,p(\boldsymbol{\psi})d\boldsymbol{\psi} \\ &= p(\mathbf{Z})\int p(\mathbf{X},\mathbf{Y}|\boldsymbol{\psi})\,p(\boldsymbol{\psi})d\boldsymbol{\psi} \\ &= p(\mathbf{X},\mathbf{Y})\,p(\mathbf{Z}) \end{aligned} \tag{10}$$

Substituting Eq. 9 and Eq. 10 (noting that $p(\boldsymbol{\theta}) = p(\mathbf{X},\mathbf{Y},\mathbf{Z})$) into Eq. 6, the posterior distribution can now be expressed as:

$$p(\boldsymbol{\psi}|\mathbf{D}) = \iint p(\mathbf{X},\mathbf{Y}|\mathbf{D})\,p(\mathbf{Y}|\mathbf{X},\boldsymbol{\psi})\,p(\boldsymbol{\psi}|\mathbf{X})\,p(\mathbf{Y}|\mathbf{X})^{-1}d\mathbf{X}d\mathbf{Y} \tag{11}$$

It is reasonable to assume that $p(\boldsymbol{\psi}|\mathbf{X})$ is slow-varying with respect to $\boldsymbol{\psi}$ compared to $p(\mathbf{Y}|\mathbf{X},\boldsymbol{\psi})$ since only knowing $\mathbf{X}$ does not provide much information about $\boldsymbol{\psi}$ due to the absence of $\mathbf{Y}$. On the other hand, in the absence of knowledge about $\boldsymbol{\psi}$ that characterises the probabilistic description of $\mathbf{Y}$ given $\mathbf{X}$, $p(\mathbf{Y}|\mathbf{X})$ is slow-varying with respect to $\mathbf{X}$ and $\mathbf{Y}$ compared to $p(\mathbf{Y}|\mathbf{X},\boldsymbol{\psi})$ and hence can be assumed practically constant. We can now express $p(\boldsymbol{\psi}|\mathbf{D})$ as

$$p(\boldsymbol{\psi}|\mathbf{D}) \propto \iint p(\mathbf{X},\mathbf{Y}|\mathbf{D})\,p(\mathbf{Y}|\mathbf{X},\boldsymbol{\psi})d\mathbf{X}d\mathbf{Y} \tag{12}$$

Eq. 12 provides a framework for incorporating the identification uncertainty of training data when inferring about the hyper parameters of the data driven model. It expresses the posterior distribution of $\boldsymbol{\psi}$ in terms of the posterior PDF of training set given the system measurements as well as the posterior PDF of output data given input training data and hyperparameters. Compared to Eq. 1, Eq. 12 is computationally tractable since the first term $p(\mathbf{X},\mathbf{Y}|\mathbf{D})$ results directly from Bayesian inference of $\{\mathbf{X},\mathbf{Y}\}$ based on measurement $\mathbf{D}$ and the second term $p(\mathbf{Y}|\mathbf{X},\boldsymbol{\psi})$ results directly from the data driven model adopted.

3. Efficient Algorithm for OMA Data using GP

The Bayesian framework in the last section is applicable in a general context as long as the assumptions are met. In this section, it is further specialized to OMA data with a GP model adopted as data driven model. The resulting algorithm is efficient for computation and applied to SHM data of a tall building in the next section.

First consider $p(\mathbf{X},\mathbf{Y}|\mathbf{D})$ in the context of OMA. For given measurement set D_i, the posterior distribution $p(x_i, y_i|D_i)$ can be inferred using a Bayesian operational modal analysis (BAYOMA) approach (Au 2012a; b). For sufficient data, modal analysis problem is 'globally identifiable' (Beck and Katafygiotis 1998). The posterior PDF $p(x_i, y_i|D_i)$ then can be approximated by a Gaussian PDF (Beck and Katafygiotis 1998):

$$p(x_i, y_i|D_i) \approx \mathcal{N}\left(\begin{bmatrix} x_i \\ y_i \end{bmatrix} \middle| \begin{bmatrix} \hat{x}_i \\ \hat{y}_i \end{bmatrix}, \hat{\mathbf{C}}_i\right) \tag{13}$$

where $\{\hat{x}_i, \hat{y}_i\}$ are the most probable value (MPV) of $\{x_i, y_i\}$ and $\hat{\mathbf{C}}_i$ is the corresponding posterior covariance matrix. The identification results of $\{x_i, y_i\}$ from D_i will be taken as training data for constructing the Gaussian process model in order to learn the relationship between $\mathbf{X}$ and $\mathbf{Y}$. Given $\{x_i, y_i\}_{i=1}^{n_s}$, the system measurements $\{D_i\}_{i=1}^{n_s}$ are assumed to be independent. Together with the fact that D_i only depends on $\{x_i, y_i\}$, we have

$$p(\mathbf{D}|\mathbf{X},\mathbf{Y}) = \prod_{i=1}^{n_s} p(D_i|\mathbf{X},\mathbf{Y}) = \prod_{i=1}^{n_s} p(D_i|x_i, y_i) \tag{14}$$

Using Baye's theorem with a flat prior on $(\mathbf{X},\mathbf{Y})$, $p(\mathbf{X},\mathbf{Y}|\mathbf{D})$ can be approximated as

$$p(\mathbf{X},\mathbf{Y}|\mathbf{D}) \approx \mathcal{N}\left(\begin{bmatrix} \mathbf{X} \\ \mathbf{Y} \end{bmatrix} \middle| \begin{bmatrix} \hat{\mathbf{X}} \\ \hat{\mathbf{Y}} \end{bmatrix}, \begin{bmatrix} \mathbf{C}_\mathbf{X} & \mathbf{C}_\mathbf{XY} \\ \mathbf{C}_\mathbf{YX} & \mathbf{C}_\mathbf{Y} \end{bmatrix}\right) \tag{15}$$

which is a Gaussian PDF with $\mathbf{C}_\mathbf{X} = diag([c_{x_1} \dots c_{x_{n_s}}])$, $\mathbf{C}_\mathbf{Y} = diag([c_{y_1} \dots c_{y_{n_s}}])$ and $\mathbf{C}_\mathbf{XY} = \mathbf{C}_\mathbf{YX} = diag([c_{x_1 y_1} \dots c_{x_{n_s} y_{n_s}}])$.

Now consider $p(\mathbf{Y}|\mathbf{X},\boldsymbol{\psi})$ when a GP model is adopted. Without loss of generality, a regression model with unknown relationship f between a given input x and output y can be written as:

$$y = f(x) + e \tag{16}$$

where e accounts for modelling error. Instead of parameterising f, a GP model assumes that given the input data $\mathbf{X}$ the output data $\mathbf{Y}$ are jointly Gaussian:

$$p(\mathbf{Y}|\mathbf{X},\boldsymbol{\psi}) \sim \mathcal{GP}(\mathbf{M}, \mathbf{K} + \sigma_e^2 \mathbf{I}) \tag{17}$$

where mean $\mathbf{M}$ and covariance $\mathbf{K}$ are functions of the input training data $\mathbf{X}$ and hyperparameters $\boldsymbol{\psi}$. The modelling error e is assumed to be Gaussian and its variance σ_e^2 is a hyper parameter as well.

To further facilitate computation, rewrite $p(\mathbf{X},\mathbf{Y}|\mathbf{D})$ in Eq. 15 as

$$p(\mathbf{X},\mathbf{Y}|\mathbf{D}) = p(\mathbf{Y}|\mathbf{X},\mathbf{D})\, p(\mathbf{X}|\mathbf{D}) \tag{18}$$

Clearly, $p(\mathbf{X}|\mathbf{D})$ is a Gaussian PDF with mean $\hat{\mathbf{X}}$ and covariance matrix $\mathbf{C}_\mathbf{X}$. On the other hand, $p(\mathbf{Y}|\mathbf{X},\mathbf{D})$ is the conditional PDF, which from standard results is also a Gaussian PDF for $\mathbf{Y}$ with mean $\hat{\mathbf{Y}} + \mathbf{C}_\mathbf{YX}\mathbf{C}_\mathbf{X}^{-1}(\mathbf{X} - \hat{\mathbf{X}})$ and covariance matrix $\mathbf{C}_\mathbf{Y} - \mathbf{C}_\mathbf{YX}\mathbf{C}_\mathbf{X}^{-1}\mathbf{C}_\mathbf{XY}$. Substituting Eq.17 and Eq. 18 into Eq. 12 gives

$$p(\boldsymbol{\psi}|\mathbf{D}) \propto \int F(\mathbf{X})\, p(\mathbf{X}|\mathbf{D})\, d\mathbf{X} \tag{19}$$

where

$$F(\mathbf{X}) = (2\pi)^{-n_s/2} \det(\mathbf{C}_\mathbf{W})^{-1/2} \exp\left(-\frac{1}{2}\mathbf{W}^T \mathbf{C}_\mathbf{W}^{-1} \mathbf{W}\right) \tag{20}$$

with

$$\mathbf{W} = \hat{\mathbf{Y}} + \mathbf{C}_\mathbf{YX}\mathbf{C}_\mathbf{X}^{-1}(\mathbf{X} - \hat{\mathbf{X}}) - \mathbf{M} \tag{21}$$

$$\mathbf{C}_\mathbf{W} = \mathbf{C}_\mathbf{Y} - \mathbf{C}_\mathbf{YX}\mathbf{C}_\mathbf{X}^{-1}\mathbf{C}_\mathbf{XY} + \mathbf{K} + \sigma_e^2 \mathbf{I} \tag{22}$$

The resulting integrand in Eq. 19 generally depends on $\mathbf{X}$ in a nonlinear manner and is not proportional to a standard probabilistic distribution. Without resorting to brute-force numerical integration that is prohibitive, a Gaussian type approximation (Girard 2004) is adopted, which gives

$$p(\boldsymbol{\psi}|\mathbf{D}) \approx (2\pi)^{-n_s/2} \det(\mathbf{C}'_\mathbf{W})^{-1/2} \exp\left(-\frac{1}{2}\mathbf{W}'^T \mathbf{C}'^{-1}_\mathbf{W} \mathbf{W}'\right) \tag{23}$$

where

$$\begin{aligned} \mathbf{W}' &= \int \mathbf{W} \mathcal{N}(\mathbf{X}|\hat{\mathbf{X}}, \mathbf{C}_\mathbf{X})\, d\mathbf{X} \\ &= \hat{\mathbf{Y}} - \mathbf{M}' \end{aligned} \tag{24}$$

$$\begin{aligned} \mathbf{C}'_\mathbf{W} &= \mathbf{C}_\mathbf{Y} + \sigma_e^2 \mathbf{I} + \mathbf{K}' \\ &- \mathbf{C}_\mathbf{XY}\mathbf{C}_\mathbf{X}^{-1} \int (\mathbf{X} - \hat{\mathbf{X}}) \mathbf{M}^T \mathcal{N}(\mathbf{X}|\hat{\mathbf{X}}, \mathbf{C}_\mathbf{X})\, d\mathbf{X} \\ &- \mathbf{C}_\mathbf{XY}\mathbf{C}_\mathbf{X}^{-1} \int \mathbf{M} (\mathbf{X} - \hat{\mathbf{X}})^T \mathcal{N}(\mathbf{X}|\hat{\mathbf{X}}, \mathbf{C}_\mathbf{X})\, d\mathbf{X} \\ &+ \int \mathbf{M}\mathbf{M}^T \mathcal{N}(\mathbf{X}|\hat{\mathbf{X}}, \mathbf{C}_\mathbf{X})\, d\mathbf{X} - \mathbf{M}'\mathbf{M}'^T \end{aligned} \tag{25}$$

with

$$\mathbf{M}' = \int \mathbf{M} \mathcal{N}(\mathbf{X}|\hat{\mathbf{X}}, \mathbf{C}_\mathbf{X})\, d\mathbf{X} \tag{26}$$

$$\mathbf{K}' = \int \mathbf{K} \mathcal{N}(\mathbf{X}|\hat{\mathbf{X}}, \mathbf{C}_\mathbf{X})\, d\mathbf{X} \tag{27}$$

Whether the analytical expressions of $\mathbf{W}'$ and $\mathbf{C}'_\mathbf{W}$ are available still depends on the form of the mean and covariance function. For a commonly used GP model with zero mean function and squared exponential covariance function, $\mathbf{W}'$ and $\mathbf{C}'_\mathbf{W}$ can be simplified as

$$\mathbf{W}' = \hat{\mathbf{Y}} \tag{28}$$

$$\mathbf{C}'_\mathbf{W} = \mathbf{C}_\mathbf{Y} + \sigma_e^2 \mathbf{I} + \mathbf{K}' \tag{29}$$

with

$$\mathbf{K}'(x_i, x_j) = \frac{\sigma_f^2 \exp\left(-\frac{1}{2}(\hat{x}_i - \hat{x}_j)^T \left(w + c_{x_i} + c_{x_j}\right)^{-1} (\hat{x}_i - \hat{x}_j)\right)}{\left|1 + w^{-1}\left(c_{x_i} + c_{x_j}\right)(1 - \delta_{ij})\right|^{1/2}} \tag{30}$$

Here δ_{ij} is the Kronecker delta, i.e., $\delta_{ij} = 1$ if $i = j$ and zero otherwise. For inferring the hyper parameters $\boldsymbol{\psi}$, it is more convenient to work with the negative log-likelihood function

$$L_{\psi} = \frac{1}{2}\ln\det\left(\mathbf{C}'_{\mathrm{W}}\right) + \frac{1}{2}\mathbf{W}'^{T}\mathbf{C}'^{-1}_{\mathrm{W}}\mathbf{W}' \quad (31)$$

such that

$$p\left(\psi\middle|\mathbf{D}\right) \propto \exp\left(-L_{\psi}\right) \quad (32)$$

The hyper parameter ψ now can be obtained by maximising $p\left(\psi\middle|\mathbf{D}\right)$, or equivalently minimising L_{ψ}.

4. Application of Tall Building SHM

The proposed method has been validated in a previous work (Zhu and Au 2020). This section focuses on further applying the proposed method to SHM of a tall building during a typhoon event, illustrating its feasibility to SHM data in real applications. The instrumented building is 320m tall and 50m by 50m in plan located in Hong Kong. Benchmark tests have been conducted under normal wind conditions with four triaxial accelerometers placed at four corners on the roof. Detailed modal identification results (namely Building B) can be found in Au et al. (2012).

The vibration response of the building was measured during Typhoon Vicente in July 2012. A triaxial force balance accelerometer was placed in a secure room on the roof of the buildings to measure vibration response. Forty-eight hours of acceleration time history data were recorded. The instrument (sensor and data acquisition unit) has a noise level of $1\mu g/\sqrt{\mathrm{Hz}}$ and the data was logged using a 24bit digital signal recorder at a sampling rate of 50Hz. The whole time history data is divided into non-overlapping segments each with a duration of 30mins. The data is modelled as stationary stochastic process within each segment and Bayesian OMA techniques is applied to identify the modal parameters. The change of dynamic properties against environmental variations are investigated. Specifically, the dynamic properties here refer to the natural frequency and damping ratio and the environmental variation refers to that of the power spectral density (PSD) of modal force (which reflects the intensity of the wind).

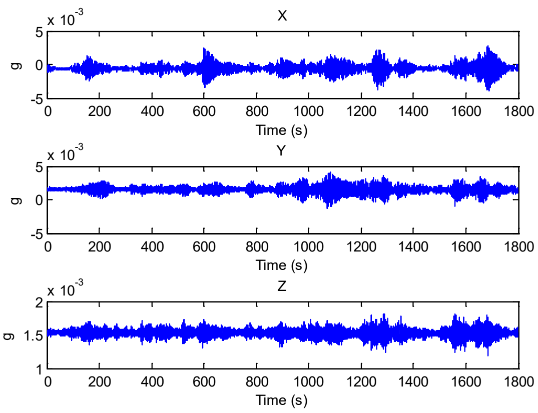

Figure 2. Time history data at 23rd July 2012 11:30pm

Fig. 2 shows a typical time history data starting at 23rd July 2012 11:30pm when the wind was strong. The maximum acceleration response of the structure is around 0.005g. Fig. 3 shows the corresponding singular value spectrum (i.e., a plot of the square root of the eigenvalues of the real part of the spectral density matrix against frequency). Modal analysis here focuses on the first two modes marked in the figure, where '[-]' denotes the selected frequency band and 'o' denotes the initial guess of natural frequency. These two modes are translational modes identified simultaneously based on the same band as they are closely spaced. Fig. 4 shows the identified mode shapes of these two modes. The interactions between these two modes also increase the identification uncertainty of the modal parameters.

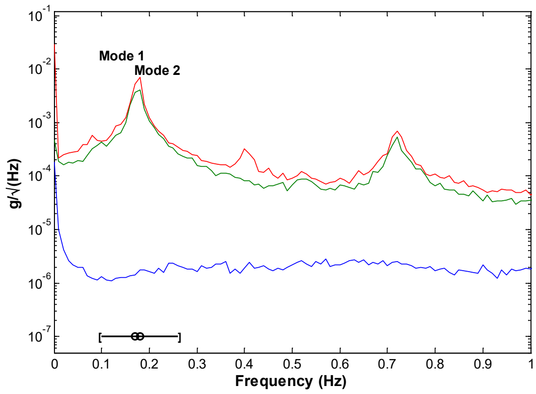

Figure 3. Singular value spectrum

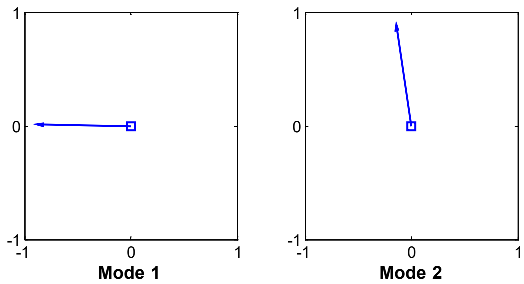

Figure 4. Identified mode shapes

Fig. 5 and Fig. 6 show the identified natural frequency against the modal force PSD for Mode 1 and Mode 2, respectively. The circle in the figure denotes the MPV and the error bar represents +/-2 posterior standard deviation. There is an inverse trend between the natural frequency and modal force PSD, indicating the amplitude dependence of the tested structure. The predictive mean values with the predictive 95% confidence bounds (i.e., +/-2 standard deviation) based on the proposed method and conventional GP model have also been plotted in the figures. The squared exponential function is selected as the covariance function and the mean function is set as zero for both models. It can be seen that the prediction from the proposed model is similar to that based on the conventional GP model based on the training data of mode 1. This is reasonable since the posterior uncertainties among the training data are similar. This is not the case for the identified damping ratio, however.

Fig. 7 and Fig. 8 show the identified damping ratio against modal force PSD with model predictions of these two methods, respectively. Discrepancies in model predictions can be found around modal force PSD of $10^{-10}\mathrm{g}^2/\mathrm{Hz}$ between the proposed method and conventional GP model for Mode 1. This is due to the large identification uncertainty of the training points marked in the red square shown in Fig. 7. The conventional GP model does not consider the uncertainty of individual training points and so it tries to fit the training points with the same weight. This is not the case for the proposed method. The training points marked in the red square take less weight when the model is trained by the proposed method due to their large uncertainty. The resulting data model has lower sensitivity to training points with large uncertainty compared to the model based on conventional GP method, in a manner consistent with the uncertainty of parameters.

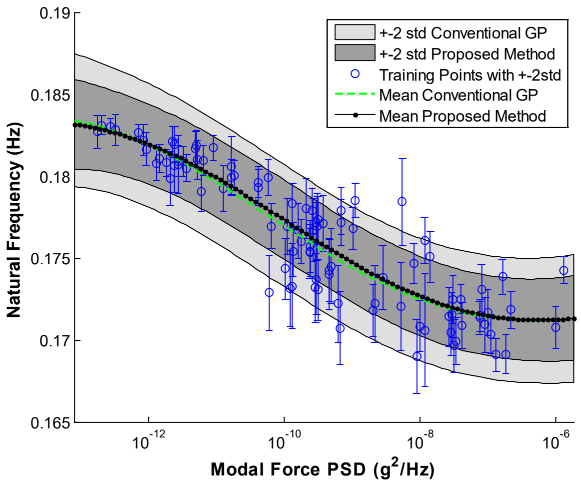

Figure 5. Identified natural frequency against modal force PSD with data driven models, Mode 1

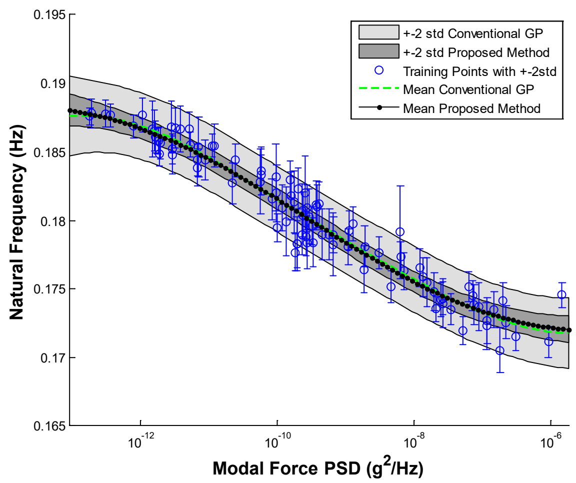

Figure 6. Identified natural frequency against modal force PSD with data driven models, Mode 2

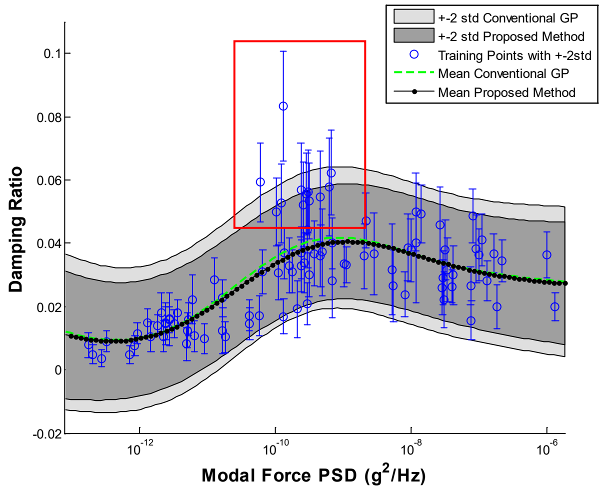

Figure 7. Identified damping ratio against modal force PSD with data driven models, Mode 1

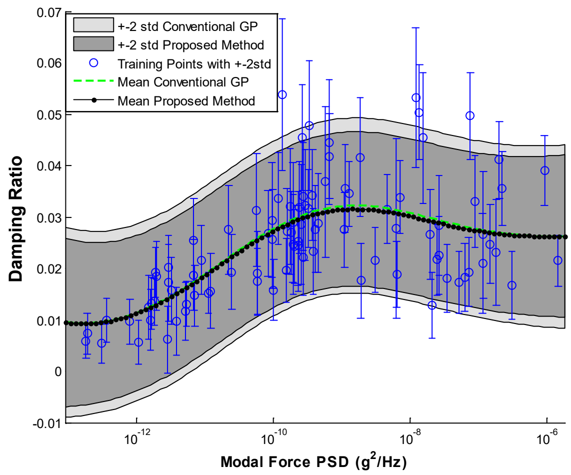

Figure 8. Identified damping ratio against modal force PSD with data driven models, Mode 2

5. Conclusions

Training data are commonly assumed as precise values without uncertainty in conventional data driven models. However, this may not be the case when applying data driven models to SHM data. The training data may not be directly obtained but identified from the measured structural response data, which inevitably carries identification uncertainties. A Bayesian data driven framework has been used in this work which incorporates the identification uncertainty rigorously when inferring the data driven model. Efficient algorithm has been presented for OMA data with GP adopted as data driven models.

The method has been applied to SHM of a tall building under a typhoon event, which illustrates its feasibility to real data. It was shown that when the variation of the identification uncertainty among the training data is small, the proposed method has similar performance compared to classic GP model which does not consider uncertainty. The proposed method has shown robustness to large discrepancies among the identification uncertainty of the training data sets. The classical GP model treat all the training data equally as the associated uncertainty of each individual training data is not considered. On the other hand, the proposed method accounts for the training points in accordance with their identification uncertainty.

Acknowledgements

This paper is supported by UK Engineering & Physical Research Council (EP/R006768/1), a major programme grant to create a robustly validated virtual prediction tool called a "digital twin". The financial support is gratefully acknowledged. The work related to this paper supports the development of fundamental verification and validation framework for digital twins of structures whose dynamics are identified from stochastic data.

References

Au, S.-K. (2012a). "Fast Bayesian ambient modal identification in the frequency domain, Part I: Posterior most probable value." *Mechanical Systems and Signal Processing*, Elsevier, 26, 60–75.

Au, S.-K. (2012b). "Fast Bayesian ambient modal identification in the frequency domain, Part II: Posterior uncertainty." *Mechanical Systems and Signal Processing*, Elsevier, 26, 76–90.

Au, S.-K. (2017). *Operational Modal Analysis: Modeling, Bayesian Inference, Uncertainty Laws*. Springer.

Au, S.-K., Zhang, F.-L., and To, P. (2012). "Field observations on modal properties of two tall buildings under strong wind." *Journal of Wind Engineering and Industrial Aerodynamics*, Elsevier, 101, 12–23.

Beck, J. L., and Katafygiotis, L. S. (1998). "Updating Models and Their Uncertainties. I: Bayesian Statistical Framework." *Journal of Engineering Mechanics*.

Chang, P. C., Flatau, A., and Liu, S. C. (2003). "Review Paper: Health Monitoring of Civil Infrastructure." *Structural Health Monitoring*, 2(3), 257–267.

Girard, A. (2004). "Approximate methods for propagation of uncertainty with Gaussian process models." Citeseer.

Kopsaftopoulos, F., Nardari, R., Li, Y. H., and Chang, F. K. (2018). "A stochastic global identification framework for aerospace structures operating under varying flight states." *Mechanical Systems and Signal Processing*.

Rasmussen, C. E., and Williams, C. K. . I. (2005). *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning). The MIT Press, Cambridge, MA, USA*.

Reynders, E., Wursten, G., and de Roeck, G. (2014). "Output-only structural health monitoring in changing environmental conditions by means of nonlinear system identification." *Structural Health Monitoring*.

Sohn, H., Farrar, C. R., Hemez, F., and Czarnecki, J. (2003). "A Review of Structural Health Monitoring Literature 1996 – 2001." *Los Alamos National Laboratory, USA*.

Spiridonakos, M. D., Chatzi, E. N., and Sudret, B. (2016). "Polynomial Chaos Expansion Models for the Monitoring of Structures under Operational Variability." *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering*.

Wenzel, H., and Pichler, D. (2005). *Ambient vibration monitoring*. Wiley, UK.

Zhu, Y. C., and Au, S. K. (2020). "Bayesian data driven model for uncertain modal properties identified from operational modal analysis." *Mechanical Systems and Signal Processing*.