A Bayesian Filtering Approach to Operational Modal Analysis with Recovery of Forcing Signals

T. J. Rogers [1], **K. Worden** [1], **G. Manson** [1], **U. T. Tygesen** [2], **E. J. Cross** [1]
[1] Department of Mechanical Engineering, University of Sheffield,
Mappin Street, Sheffield, S1 3JD, UK
e-mail: **trogers3@sheffield.ac.uk**

[2] Ramboll Energy, Bavnehøjvej 5, DK-6700 Esbjerg Denmark

Abstract
The problem of Operational Modal Analysis (OMA) is a difficult and active area of research; this is especially the case when loads on a structure, which can't be measured, are harmonic and close to the resonance frequencies of the structure, and if the structure is lightly damped. This situation can arise in many areas of engineering, one key example is in the offshore energy industry where wave loads are commonly a narrowband random loading and it is not easy to structurally alter the system so that none of the resonance frequencies coincide with the dominant frequency of the loading. This paper presents a *grey-box* approach to the OMA problem, where a physical model of the system dynamics is augmented using a technique from machine learning. A Latent Force Model (LFM) is used to represent the loading acting on a structure, where the forcing is assumed to be a Gaussian Process. The LFM is transformed into a state-space formulation, as are the structural dynamics, and inference is performed simultaneously over the system parameters and the hyperparameters of the Gaussian Process via a Markov-Chain Monte-Carlo method. This is shown to give realistic distributions over the system parameters and produce a good prediction of the latent forces acting on the system.

1 Introduction

Experimental modal analysis of linear systems is a well-established area of interest within structural dynamics. Techniques for inferring the system parameters of linear systems where input and output data are present are now mature and can be found in many good undergraduate textbooks, for example see [1]. In a wider sense, the handling of second-order linear time-invariant (LTI) systems, of which physical dynamic systems are a subset, is also well covered in the literature [2]. Unfortunately, for many structures which are encountered by engineers, it is not possible to gather information regarding the input to the system — the forcing. This is due to two important factors; first, it is not possible to place sensors in the load path on some structures such as buildings undergoing base excitation from earthquakes. Secondly, it can be prohibitively expensive to design and fit sensors which can withstand the environmental conditions in which the structure operates. This is the case in the offshore energy industry where wave loading represents a phenomenon which is difficult to measure and which produces an environment in which sensors must be very robust.

The current solution to this is to use techniques in Operational Modal Analysis (OMA), or output-only modal analysis. Reviews of some techniques which may be used can be found in [3, 4]; however, a common problem encountered is the identification of modal behaviour when the loading has significant harmonic components which are close to the modes of the structure and when the structure is lightly damped. In this case it becomes very difficult to separate the narrow peaks in the frequency domain due to the harmonic loads and those from the resonances of the structure.

This paper makes use of a state-space form of the system, which is a common approach in OMA techniques, and one that has been shown to be effective for state/input estimation [5, 6]. It has become commonplace to use the Extended Kalman Filter (EKF) formulation where the parameters of the model are included as additional states. In this way recursive estimates of the system parameters can be made within the Bayesian filtering setting [7]. Lourens *et. al* [8] show a Kalman filter-based formulation for estimation of the states of the structure and the forcing signal based on fixed parameter sets from a reduced order finite element (FE) model.

This paper also aims to incorporate a novel use of the machine learning Latent Force Model (LFM) [9] in a state space formulation [10], to allow efficient inference with a nonparametric prior over the forcing function. Alongside this, Bayesian estimates of the system properties can be obtained using MCMC inference over the state space model; this avoids the linearisation present in the EKF formulation for parameter estimates. Being a Bayesian model of the full system, this methodology returns distributions over the parameters and also the forcing function which give indications of the uncertainty in both the inputs and the parameters. This uncertainty can be carried forward into further quantification analysis or used for heuristic assessment of model quality.

The layout of the paper is as follows: the theoretical foundations of the model are laid out in Sections 2 to 4, covering an introduction to Gaussian Processes, state-space modelling, and the Latent Force Model. The application of this model to the problem of OMA is discussed in Section 5, including the need to constrain the model with informative prior distributions. Section 6 presents results from use of this procedure on a simulated dataset, and conclusions are made in Section 7.

2 Gaussian Process Regression

Gaussian Processes (GPs) [11, 12] are a popular nonparametric Bayesian machine learning technique. The most interpretable explanation is as a distribution over functions, so each random draw from a GP is a function from a functional family defined by the covariance function and specified by the data that has been used to train the process. In this section we adopt notation from within the GP community which is in line with that in [12], and which presents the general case for any GP, in the rest of the paper the notation will be consistent with the usual practice in state-space models for clarity.

A Gaussian Process is fully defined in terms of its mean function $m(\boldsymbol{x})$ (the mean is commonly assumed to be zero and this convention is adopted here although addition of an explicit mean function is trivial) and its covariance function $k(\boldsymbol{x}, \boldsymbol{x}')$, see Eq. (1). The covariance function allows the computation of the covariance between any two data points $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ such that the element in the covariance matrix K_{ij} is given by $k(\boldsymbol{x}_i, \boldsymbol{x}_j)$. This allows the covariance matrix for any finite number of data points to be calculated analytically. Thus,

$$f(\mathbf{x}) \sim \mathcal{GP}\left(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')\right) \tag{1}$$

Assuming there is a finite set of observed training data $\mathcal{D}$ with N observations of input-output pairs: $\mathcal{D} = \{\boldsymbol{x}_i, y_i\}_{i=1}^{N}$ where $y_i = f(\boldsymbol{x}_i) + \varepsilon$, $\varepsilon \sim \mathcal{N}\left(0\,,\ \sigma_n^2\right)$. This equation states that there is some underlying functional relationship $f(\boldsymbol{x})$ where the observations y are corrupted by white Gaussian noise. In this case, the GP distribution collapses to a Multivariate Normal distribution in N dimensions,

$$\boldsymbol{y} \sim \mathcal{N}\left(\mathbf{0}\,,\ K\left(X, X\right) + \sigma_n^2 \mathbb{I}\right) \tag{2}$$

Where $\boldsymbol{y}$ is the vector of all observed noisy outputs and X the matrix of all observed inputs in $\mathcal{D}$. When a new input point $\boldsymbol{x}^\star$ or number of input points $X^\star$ is observed, the covariance with the training data can be calculated as a joint Gaussian formed between the training and test data,

$$
\begin{bmatrix} \boldsymbol{y} \\ \boldsymbol{f}_\star \end{bmatrix} \sim \left(\mathbf{0}, \begin{bmatrix} K(X,X) + \sigma_n^2 \mathbb{I} & K(X, X^\star) \\ K(X^\star, X) & K(X^\star, X^\star) \end{bmatrix} \right). \tag{3}
$$

The predictive distribution over $\boldsymbol{f}^\star$ can be computed by considering the conditional distribution $p\left(\boldsymbol{f}^\star \mid \boldsymbol{x}^\star, X, \boldsymbol{y}\right)$. Since the joint distribution, is Gaussian the conditional distribution can be computed analytically, where it itself is a Gaussian, and the equations for the predictive mean and variance are given by,

$$
\boldsymbol{f}^\star \sim \mathcal{N}\left(\mathbb{E}\left[\boldsymbol{f}^\star\right], \ \mathbb{V}\left[\boldsymbol{f}^\star\right]\right) \tag{4a}
$$

$$
\boldsymbol{f}^\star = k(\mathbf{x}^\star, X) \left[K(X,X) + \sigma_n^2 \mathbb{I}\right]^{-1} \mathbf{y} \tag{4b}
$$

$$
\mathbb{V}\left[\boldsymbol{f}^\star\right] = k(\mathbf{x}^\star, \mathbf{x}^\star) - k(\mathbf{x}^\star, X) \left[K(X,X) + \sigma_n^2 \mathbb{I}\right]^{-1} k(X, \mathbf{x}^\star) \tag{4c}
$$

3 Bayesian State-Space Models

This paper focuses on linear Gaussian state-space models. The notation, from this section onwards, is $\boldsymbol{x}(t)$ is used to represent the state vector, and $\boldsymbol{y}(t)$ the observations in the model for the continuous case or $\boldsymbol{x}_t$ and $\boldsymbol{y}_t$ for the discrete time case. $\boldsymbol{u}(t)$ and its discrete time counterpart $\boldsymbol{u}_t$ is a known input to the system at time t. It is normally most convenient to form a continuous time state-space model since differential equations readily convert into this form. The continuous time state-space transition is defined as,

$$
\frac{\mathrm{d}\boldsymbol{x}(t)}{\mathrm{d}t} = F\boldsymbol{x}(t) + G\boldsymbol{u}(t) + Lw(t) \tag{5}
$$

Here, F is the continuous state transition matrix, G the continuous input matrix, and L the continuous time process noise matrix when $w(t)$ is a white noise process with spectral density q. Since observations of physical systems happen in discrete time, it is usual that analysis is carried out with a discrete model; the observation model is added at this point.

$$
\boldsymbol{x}_{t+1} = A\boldsymbol{x}_t + B\boldsymbol{u}_t + \boldsymbol{w}_t, \tag{6a}
$$

$$
\boldsymbol{y}_t = C\boldsymbol{x}_t + \boldsymbol{v}_t \tag{6b}
$$

with A being the discrete-time transition matrix and B the discrete-time input matrix. The process noise $\boldsymbol{w}_t$ is distributed $\mathcal{N}\left(\mathbf{0}, \ R\right)$. C is the observation matrix and the observation noise $\boldsymbol{v}_t \sim \mathcal{N}\left(\mathbf{0}, \ R\right)$. Since the solutions to models of this form are in discrete time it is necessary to convert a continuous time transition model into its corresponding discrete time representation, i.e. find A and B from F and G. This procedure is well documented; for example, see [2], and is shown here for completeness.

$$
A = \exp_m\left(F\,\Delta t\right) \tag{7a}
$$

$$
B = F^{-1}\left[\left(A - \mathbb{I}\right)G\right] \tag{7b}
$$

$\exp_m$ is the matrix exponential operator; Δt is the sampling period when discretising the continuous-time model. The continuous-time noise process $Lw(t)$ must also be discretised to perform inference over the model.

$$Q = \int_0^{\Delta t} \Phi\left(\Delta t - \tau\right) L q L^{\mathsf{T}} \Phi\left(\Delta t - \tau\right) \mathrm{d}\tau \tag{8}$$

where, $\Phi(\tau) = \exp_m\left(F\tau\right)$. Once discretised, it is now possible to recover the *filtering distributions* $p\left(\boldsymbol{x}_t \mid \boldsymbol{y}_{1:t}\right)$ and the *smoothing distributions* $p\left(\boldsymbol{x}_t \mid \boldsymbol{y}_{1:T}\right)$, for a time point t in a data record that is T time points long, using the Kalman filtering equations [13] and the Rauch-Tung-Striebel (RTS) smoother [14].

4 The Latent Force Approach

The Latent Force Model (LFM) [9, 15], is an approach for combining GPs with a mechanistic model where the GP is used to represent the forcing experienced by the system. In this way, the output of the GP can include modelling of the dynamics of a system. This model is a form of *grey-box*, where known physical processes are augmented with machine learning technology to improve model performance. Of most interest, from the point of view of structural dynamics is a second order LFM which has the familiar form of a P degree of freedom linear dynamical system,

$$M\ddot{\boldsymbol{x}} + C\dot{\boldsymbol{x}} + K\boldsymbol{x} = F \tag{9}$$

Additionally, in the LFM, F is defined as a set of independent Gaussian Processes in time,

$$\begin{aligned} p\left(F \mid \boldsymbol{t}\right) &= \prod_{p=1}^{P} \boldsymbol{f}_p = \prod_{p=1}^{P} \mathcal{N}\left(\mathbf{0},\ K\left(\boldsymbol{t}_p, \boldsymbol{t}_p\right)\right) \\ \boldsymbol{f}_p &\sim \mathcal{GP}\left(\mathbf{0}, k(t, t')\right) \end{aligned} \tag{10}$$

where the covariance of each of the latent forces in time is defined by the covariance function $k(t, t')$ giving rise to the covariance matrix $K_{\boldsymbol{f}_p, \boldsymbol{f}_p}$ which is the covariance between the latent forces at all points in time $K\left(\boldsymbol{t}_p, \boldsymbol{t}'_p\right)$ with $\boldsymbol{t}_p$ being the vector of all time points when the output is observed. While this model has produced some very promising results in modelling, the main challenge in implementing this form of model is its computational complexity, which is $\mathcal{O}\left(N^3 P^3\right)$, where N here is the length of the signal in time. This issue can be addressed using sparse methodologies that have become popular within the GP literature, good reviews can be found in [16, 17], more specifically, [18] shows how a multiple output GP, such as the one in the LFM can be computed in a sparse manner, reducing the computational complexity from $\mathcal{O}\left(N^3 R^3\right)$ to $\mathcal{O}\left(N^3 R\right)$ using the partially independent training conditional (PITC) sparse approximation.

However, work by Hartikainen and Särkkä [19] has shown that, for GPs where the input is temporal and the covariance function is stationary, the inference procedure can be converted into a linear time-invariant (LTI) stochastic differential equation (SDE) which has a state-space formulation. This state space model can be solved exactly, for certain cases, using a Kalman filter [13] and Rauch-Tung-Striebel (RTS) smoother [14]. This methodology computes in $\mathcal{O}\left(NP^3\right)$ time which is a significant improvement given that, usually, $N \gg P$.

This is the case if a GP is defined to have a Matérn covariance function [20] between any two points in time. Since it is a stationary covariance, it is defined in terms of the absolute difference between the two input points, $r = |t - t'|$. The Matérn class of covariance functions are governed by a smoothness parameter which gives rise to a number of special cases. Setting $\nu = 1/2$ gives rise to the exponential covariance function which is equivalent to an Ornstein-Uhlenbeck process, and as $\nu \to \infty$, the popular Squared Exponential (SE) covariance function [12], which is also referred to as the Gaussian kernel, is recovered. Choosing values of $\nu = p + 1/2$ for p as any non-negative integer, leads to simple expressions for Matérn covariance functions. In the case with smoothness parameter $\nu = 3/2$, the covariance function can be written as,

$$k(t,t') = \left(1 + \frac{\sqrt{3}r}{\ell}\right) \exp\left(-\frac{\sqrt{3}r}{\ell}\right) \tag{11}$$

Following Hartikainen and Särkkä [19] and converting the covariance function into an LTI SDE requires the spectral density, by taking the Fourier transform. Setting $\lambda = \sqrt{2\nu}\big/\ell$ and $\nu = p + 1/2$, in the Matérn class of covariance functions it can be shown that,

$$S(\omega) \propto \left(\lambda^2 + \omega^2\right)^{-(p+1)}. \tag{12}$$

This can be factorised to recover the transfer function needed to convert to a state-space model, and the stable part of that transfer function can be written down,

$$H\left(\mathrm{i}\omega\right) = \left(\lambda + \mathrm{i}\omega\right)^{-(p+1)}. \tag{13}$$

An LTI SDE of order m has the form shown in Eq. (14), where $f(t)$ is a random process of interest, $\boldsymbol{a}$ is some set of coefficients, and $w(t)$ is a white noise process which has a spectral density $S_w(\omega) = q$,

$$\frac{\mathrm{d}^m f(t)}{\mathrm{d}t^m} + a_{m-1}\frac{\mathrm{d}^{m-1} f(t)}{\mathrm{d}t^{m-1}} + \ldots + a_1\frac{\mathrm{d}f(t)}{\mathrm{d}t} + a_0 f(t) = w(t) \tag{14}$$

Rearranging this into a state-space model is possible by constructing a vector $\boldsymbol{f(t)} = \left[f(t)\frac{\mathrm{d}f(t)}{\mathrm{d}t}\ldots\frac{\mathrm{d}^{m-1}f(t)}{\mathrm{d}t^{m-1}}\right]$ which gives rise to a first-order vector Markov process,

$$\frac{\mathrm{d}\boldsymbol{f(t)}}{\mathrm{d}t} = F\boldsymbol{f(t)} + Lw(t) \tag{15}$$

where,

$$F = \begin{pmatrix} 0 & 1 & & \\ & \ddots & \ddots & \\ & & 0 & 1 \\ -a_0 & \ldots & -a_{m-2} & -a_{m-1} \end{pmatrix}, \qquad L = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix} \tag{16}$$

The coefficients $a_0, \ldots, a_{m-1}$ (Eq. (16)) are the coefficients of the polynomial in the denominator of the transfer function computed from the coefficients of the polynomial expansion of $H(\mathrm{i}\omega)$, $h_0, \ldots, h_{m-1}$. The spectral density of the white noise process $S_w(\omega) = q$ is equal to the constant from $S(\omega)$ (Eq. (12)), this is,

$$q = \frac{2\,\sigma_f^2\,\pi^{1/2}\lambda^{(2p+1)}\Gamma\left(p+1\right)}{\Gamma\left(p+1/2\right)} \tag{17}$$

For the case of the Matérn covariance with $\nu = 3/2$ as in Eq. (11) the equivalent state-space model is,

$$\frac{\mathrm{d}\boldsymbol{f(t)}}{\mathrm{d}t} = \begin{pmatrix} 0 & 1 \\ -\lambda^2 & -2\lambda \end{pmatrix} \boldsymbol{f(t)} + \begin{pmatrix} 0 \\ 1 \end{pmatrix} w(t) \tag{18}$$

It is now possible to convert this continuous-time state-space model to a discrete time form and solve with a Kalman filter and RTS-smoother [2]. Since the LFM model contains a temporal Gaussian process as the latent function, and the mechanistic portion of the model has a simple state-space representation, it is natural to form a state-space representation of the LFM and this was shown in [10].

In this paper, the methodology is shown for a single degree of freedom (SDOF) oscillator as the mechanistic model, although it can be easily extended to the multi-degree of freedom (MDOF) case — or indeed systems with different order ordinary differential equations (ODEs). First, the equation of motion is converted to its state-space form, which is again a first-order vector Markov process.

Taking the SDOF equation of motion,

$$m\ddot{x} + c\dot{x} + kx = F \tag{19}$$

dividing through by m, and setting $x_1 = x$, $x_2 = \frac{\mathrm{d}x_1}{\mathrm{d}t} = \dot{x}$ in a state vector $\boldsymbol{x}$, recovers a state-space representation of the oscillator:

$$\begin{bmatrix} \dot{x} \\ \ddot{x} \end{bmatrix} = \begin{pmatrix} 0 & 1 \\ -\frac{k}{m} & -\frac{c}{m} \end{pmatrix} \begin{bmatrix} x \\ \dot{x} \end{bmatrix} + \begin{pmatrix} 0 \\ \frac{1}{m} \end{pmatrix} F \tag{20}$$

The formulation shown in Eq. (20) is useful where there are known external inputs to the system F and the effect of the forcing on the system is through the input matrix $B = [0,\ -1/m]^{\mathsf{T}}$. In the case of OMA these inputs are unknown but can be considered as one or more additional latent states in the model. When the transition for f is linear and described by the matrix F_f this system would be written,

$$\begin{bmatrix} \dot{x} \\ \ddot{x} \\ \dot{\boldsymbol{f}} \end{bmatrix} = \begin{pmatrix} 0 & 1 & 0 \\ -\frac{k}{m} & -\frac{c}{m} & \frac{1}{m} \\ 0 & 0 & F_f \end{pmatrix} \begin{bmatrix} x \\ \dot{x} \\ \boldsymbol{f} \end{bmatrix} + Lw(t) \tag{21}$$

The number of additional states that are introduced is dependent upon the model chosen for the transition of f, which is now also in first-order vector Markov process form. When using the LFM, it is clear from Eqs. (13) and (16) that the choice of covariance function for the GP chosen to represent F will affect the number of additional states required. Stein [21] suggests that Matérn kernels are a more appropriate choice of covariance function, since the Squared Exponential imposes an unrealistic smoothness assumption. A choice of $\nu = 3/2$ as in Eq. (11), is a common choice which appears to perform well, this leads to the addition of two states to the model as in Eq. (18). This procedure leads to the full four-dimensional state space model,

$$\begin{bmatrix} \dot{x} \\ \ddot{x} \\ \dot{f} \\ \ddot{f} \end{bmatrix} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ -\frac{k}{m} & -\frac{c}{m} & \frac{1}{m} & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -\lambda^2 & -2\lambda \end{pmatrix} \begin{bmatrix} x \\ \dot{x} \\ f \\ \dot{f} \end{bmatrix} + Lw(t) \tag{22}$$

The model, including the noise $Lw(t)$, must now be discretised to perform inference.

At this point, any observation model for the observed variable y can be adopted in the process; for instance, if measuring displacement it is simply Eq. (23a), or if measuring acceleration it is Eq. (23b). These are discrete-time observations, therefore, the observation matrices are multiplied by the discrete-time state values. If other observation variables were available these could also be incorporated if a suitable observation model could be formed. One has,

$$y = \begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_t \\ \dot{x}_t \\ f_t \\ \dot{f}_t \end{bmatrix} + \sigma_n^2. \tag{23a}$$

$$y = \begin{bmatrix} -\frac{k}{m} & -\frac{c}{m} & \frac{1}{m} & 0 \end{bmatrix} \begin{bmatrix} x_t \\ \dot{x}_t \\ f_t \\ \dot{f}_t \end{bmatrix} + \sigma_n^2. \tag{23b}$$

Now, in possession of a model which can represent the system, attention turns to the problem of inference. While methods such as expectation maximisation exist for this form of model [22], these maximum likelihood solutions do not give information about the parameter uncertainties in the model. Instead, in this paper, a fully Bayesian solution is adopted using Markov Chain Monte Carlo (MCMC) to perform inference over the system parameters and the GP hyperparameters. MCMC is an inference method with guaranteed convergence to the true posterior distribution of interest in the limit as the length of the chain tends to infinity [23]. The parameter vector for this model, therefore, is $\boldsymbol{\theta} = \left[m, k, c, \sigma_f^2, \ell\right]$. It is assumed that the process noise in the dynamics is small (the system is well described by a linear dynamic model) and that the observation noise is known.

The likelihood of the parameters in the model given the data observed for T time points, $p\left(\boldsymbol{\theta} \mid \boldsymbol{y}_{1:T}\right)$, is related to a quantity called the energy function $\varphi_T\left(\boldsymbol{\theta}\right)$ through a proportionality relationship.

$$p\left(\boldsymbol{\theta} \mid \boldsymbol{y}_{1:T}\right) \propto \exp\left(-\varphi_T\left(\boldsymbol{\theta}\right)\right). \tag{24}$$

The energy function $\varphi_T\left(\boldsymbol{\theta}\right)$ can be well approximated by the following recursion as the filter runs [24],

$$\varphi_t\left(\boldsymbol{\theta}\right) \simeq \varphi_{t-1}\left(\boldsymbol{\theta}\right) + \frac{1}{2}\log|2\pi S_t(\boldsymbol{\theta})| + \frac{1}{2}\boldsymbol{v}_t^{\mathsf{T}} S_t^{-1}\boldsymbol{v}_t. \tag{25}$$

$\boldsymbol{v}_t$ and S_t are defined as,

$$\boldsymbol{v}_t = \boldsymbol{y}_t - C\left(A\boldsymbol{x}_{t-1}\right), \tag{26a}$$

$$S_t = C\left(AP_{t-1}A^{\mathsf{T}} + Q\right)C^{\mathsf{T}} + R, \tag{26b}$$

when P_{t-1} is the state covariance at $t-1$. The recursion for φ_T is started at $\varphi_0 = -\log p\left(\boldsymbol{\theta}\right)$. Inference can then be performed using the standard MCMC Metropolis random-walk algorithm, which for this problem is shown in Algorithm 1.

Algorithm 1 MCMC Metropolis Random Walk for Parameter Inference in State-Space LFM

```
n_b, n_s ← Length of burn-in and desired number of samples
θ_0 ← {m_0, k_0, c_0, σ²_f0, ℓ_0}                 ▷ Set Start Points for Markov Chain
n_a ← 0, k ← 0                                     ▷ Number of accepted θ, Number of Steps
p(θ' | θ_k) = N(0, Σ_p)                           ▷ Proposal is random walk with diagonal matrix Σ_p
p(θ | y) ∝ exp{−φ_T(θ)}                           ▷ Posterior Likelihood is Proportional to the Energy Function
while k < (n_b + n_s) do
    Sample θ' from p(θ' | θ_k)
    Calculate φ_T(θ')                              ▷ Eq. (25)
    log(α) = min{−φ_T(θ') + φ_T(θ_k), 0}           ▷ Log Acceptance Ratio
    θ_{k+1} ← { θ',   n_a++   with probability α
              { θ_k,          otherwise
    k ← k + 1
end while
```

In this way, it is possible to generate samples from the true posterior of the parameter vector $\boldsymbol{\theta}$. These samples can be used for further uncertainty quantification steps or can be used for model scrutiny. The strength of this method, alongside recovering Bayesian posteriors over the system properties, is to recover, as a latent state of the system, the time series of the forcing signal applied to the system.

5 Application to Operational Modal Analysis

It is clear from the formulation in Section 4 that the LFM model is directly applicable to the problem of modal analysis. One of the strengths of the model presented here is the flexibility available due to the nonparametric form of the forcing function, as defined by the GP. This flexibility also presents a problem in the practical application of the model. As well as the normal system identification problems with the scaling of the mass, stiffness, and damping parameters which lead to non-identifiability in the model when the forcing level is unknown; the ability of the GP to model behaviour very similar to the dynamics (it is not restricted to a Gaussian white noise form) means that it can mask dynamics in the system. For this reason, it is necessary to place prior distributions over the system parameters. Prior distributions can also be placed over the hyperparameters of the GP model to control their behaviour without breaking the Bayesian paradigm within which the model has been set up.

In the current work, it is necessary to constrain at least one of the system parameters to a fixed value to resolve these problems. In this case, it is assumed that one of the system parameters is known with certainty *a priori*; this can be interpreted as a delta function prior on that parameter. The other prior distributions can be elicited from knowledge of the system; this could be from a finite element (FE) model, where if necessary, reduced-order modelling techniques can be employed. It should be noted, that when using MCMC for inference, there is no restriction on the distribution of these priors, therefore, empirical distributions can be used.

The choice of prior over the hyperparameters of the GP poses a more difficult problem as they are less interpretable in a physical sense. It is possible to set a formal prior, for example $p\left(\sigma_f^2\right) = 1$ which is a form of improper prior. However, it can be helpful to set the prior over the signal variance σ_f^2 around the expected variance of the loading and the length-scale ℓ based upon the expected frequency content of the loading.

6 Results

The results presented in this paper are from a simulated SDOF system response to a measured wave force input, where the parameters of the model are chosen such that the natural frequency of the system is close to the narrowband loading frequency. The wave loading data were measured as a part of the Christchurch Bay project in the UK [25]. This case represents the initial work into the application of LFM state-space models for OMA.

The parameters of the SDOF model are chosen to be: $m = 2000$ kg, $k = 100$ N m^{-1} and $c = 5$. These values correspond to modal parameters $\omega_n = 0.2$ Hz and $\zeta = 0.0079$. This presents a problem that is normally difficult in a OMA setting, where the loading frequency is close to the resonance frequency of a lightly damped mode.

All of the parameters of the model must take only positive values for them to be valid; therefore, the random walk of the Markov-Chain must take place in a transformed space which ensures this. The values in the transformed space are denoted $(\hat{\cdot})$, with the transformation defined as,

$$\theta = \log\left\{\exp\left(\hat{\theta}\right) + 1\right\} \tag{27}$$

where θ is the parameter of interest. Prior probability distributions over the system parameters are chosen with a random perturbation on the mean values in the system parameters to simulate priors elicited from an incorrect model (e.g. an uncalibrated FE model). To constrain the problem it is assumed that the mass of the system is known *a priori* and as such this is modelled with a delta prior for all tests.

Inference was performed as in Algorithm 1 with the burn-in set to 1×10^3 samples and the total number of samples as 1×10^5. Since the distributions may be non-Gaussian, and the distributions over the Gaussian

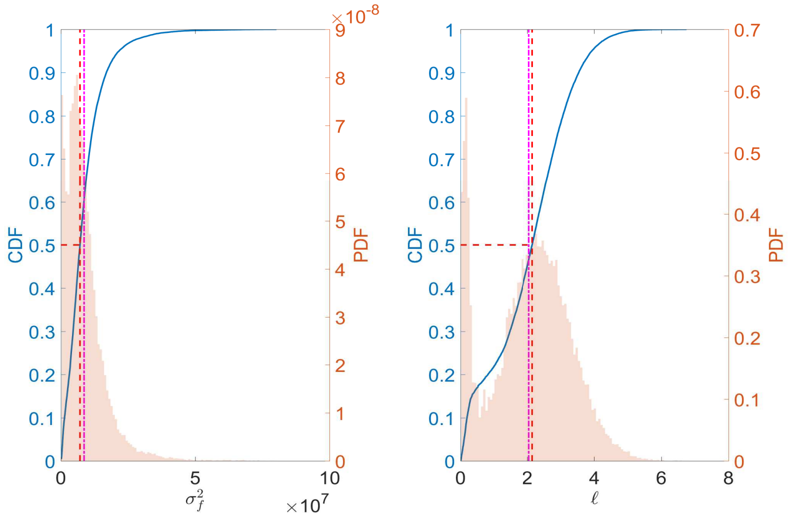

Figure 1: Plots showing the histograms and CDFs for the parameters being inferred when system parameters are known. The mean values are shown with the dot-dash ($\cdot$—) line in magenta, and the values estimated from the empirical CDF with the dashed (−−) line in red.

Parameter	True Value	Prior Mean	Posterior Mean	CDF Estimate
σ_f^2*	-	-	8.64×10^6	7.12×10^6
ℓ*	-	0.693	1.760	2.148

Table 1: Table showing results for the GP hyperparameters when the system parameters are known and fixed. (* denotes that the true values for these parameters are unknown.)

Process hyperparameters are expected to be multimodal, the parameter estimates used to calculate the example force are taken to be the values when the empirical cumulative density function (CDF) is equal to 0.5.

6.1 LFM With Known System Parameters

As an initial test the model is run with the system parameters fixed at the true values to demonstrate the ability of the LFM to model the input to the system. The priors for this case are summarised below,

$$p\left(\hat{\sigma}_f^2\right) = 1, \tag{28a}$$
$$p\left(\hat{\ell}\right) = \mathcal{N}\left(0\,,\ 2\right), \tag{28b}$$
$$p\left(m\right) = \delta\left(2000\right), \tag{28c}$$
$$p\left(k\right) = \delta\left(100\right), \tag{28d}$$
$$p\left(c\right) = \delta\left(5\right). \tag{28e}$$

Using this procedure, samples from the posteriors over the signal variance and length-scale hyperparameters were drawn. Plots of the PDF normalised histogram and CDFs are shown in Fig. 1, along with mean estimates and CDF estimates of the values. A summary of the results is shown in Table 1. In GP modelling of the latent force on the system, the signal variance of the forcing σ_f^2 is assumed unknown and a constant prior is used, the length-scale parameter has a prior to discourage very long or short length-scales. Long length-scales will cause the function in the forcing to become very smooth, the extreme of which is that it remains at the mean,

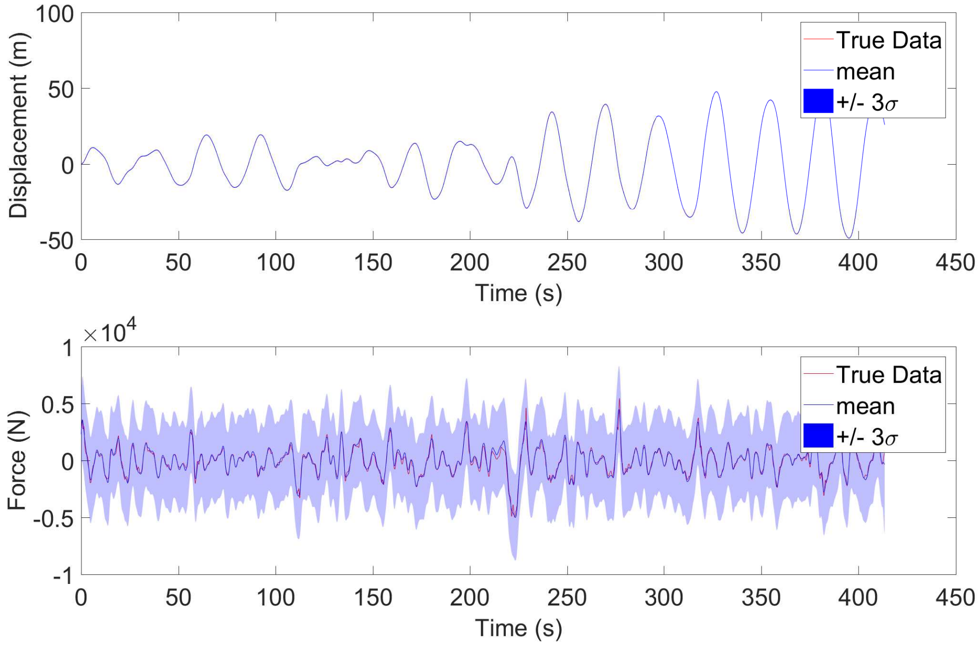

Figure 2: Smoother estimate of displacement and forcing signal on the system when m, k, and c are known. True values are shown in red, mean estimates as a solid blue line and three standard deviation confidence in the shaded regions

in this case the signal variance increases and the solution returned is one in which all the forcing is considered noise. This, although a valid solution, is not helpful in the setting of modal analysis where retrieval of the forcing time series adds a significant amount of valuable information. If the length-scale is allowed to become very small, then the GP will attempt to model the noise in the signal as if it were its own process, this is also clearly not a desirable situation. Since the distributions are skewed and multi-modal, the mean estimates are not a good value to take and instead the CDF estimate is used to predict the forcing experienced by the structure. Figure 2 shows the smoother prediction over the forcing as well as the observed state. As expected the prediction over the displacements is modelled well with very small variance. Visually the prediction over the forcing is good, to quantify the quality of the mean fit a normalised mean-squared error (NMSE) is calculated to be 3.3%. This NMSE is calculated as,

$$NMSE = \frac{100}{N\sigma_y^2} \sum_{i=1}^{N} (y_i^\star - y_i)^2 \tag{29}$$

if y is the measured value and $y^\star$ the predicted value. However, one of the strengths of this model is that in addition to this there are also confidence intervals which represent the uncertainty in the forcing.

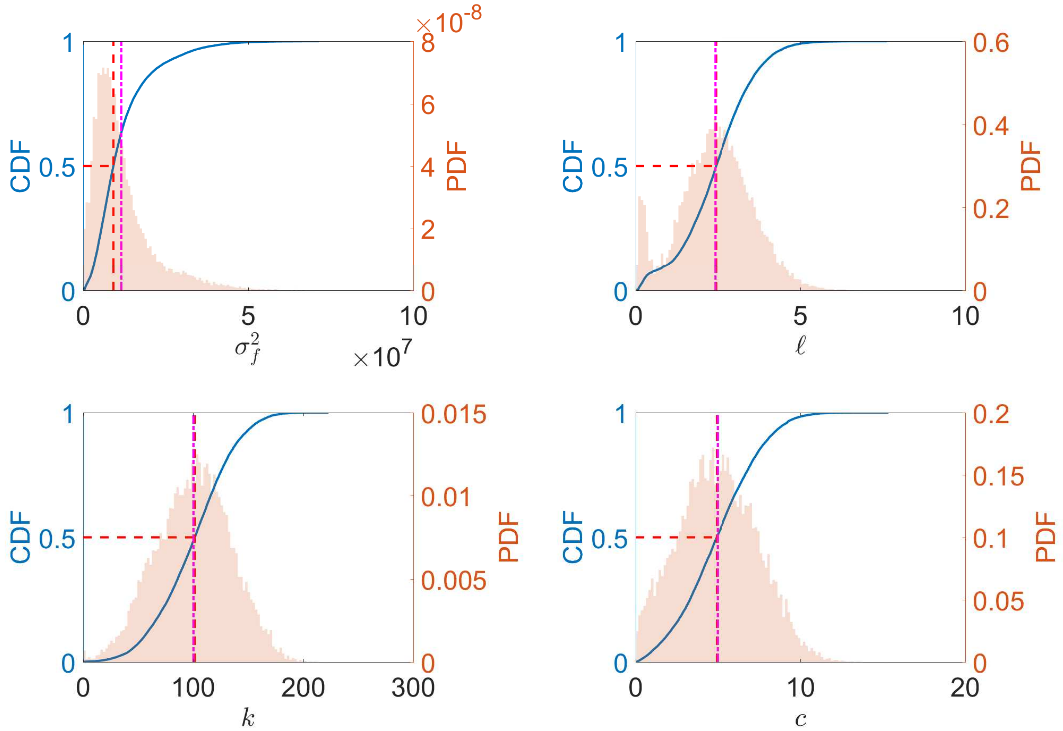

Figure 3: Plots showing the histograms and CDFs for the parameters being inferred in the full LFM OMA procedure. The mean values are shown with the dot-dash (·—) line in magenta, and the values estimated from the empirical CDF with the dashed (−−) line in red.

Parameter	True Value	Prior Mean	Posterior Mean	CDF Estimate
σ_f^2*	-	-	1.153×10^7	9.154×10^6
ℓ*	-	0.693	2.283	2.443
k	100	105.9 (5.9%)	100.2 (0.2%)	101.4 (1.4%)
c	5	4.81 (-3.8%)	4.91 (-1.8%)	4.93 (-1.4%)

Table 2: Table showing results from the full LFM OMA model, values in brackets denote the percentage error versus the true values. (* denotes that the true values for these parameters are unknown.)

6.2 OMA With The LFM

For the case of performing OMA with the GP LFM, the priors used in this model are shown below:

$$p\left(\hat{\sigma}_f^2\right) = 1, \tag{30a}$$
$$p\left(\hat{\ell}\right) = \mathcal{N}\left(0\,,\ 2\right), \tag{30b}$$
$$p\left(m\right) = \delta\left(2000\right), \tag{30c}$$
$$p\left(\hat{k}\right) = \mathcal{N}\left(105.9\,,\ 2500\right), \tag{30d}$$
$$p\left(\hat{c}\right) = \mathcal{N}\left(4.8\,,\ 6.25\right). \tag{30e}$$

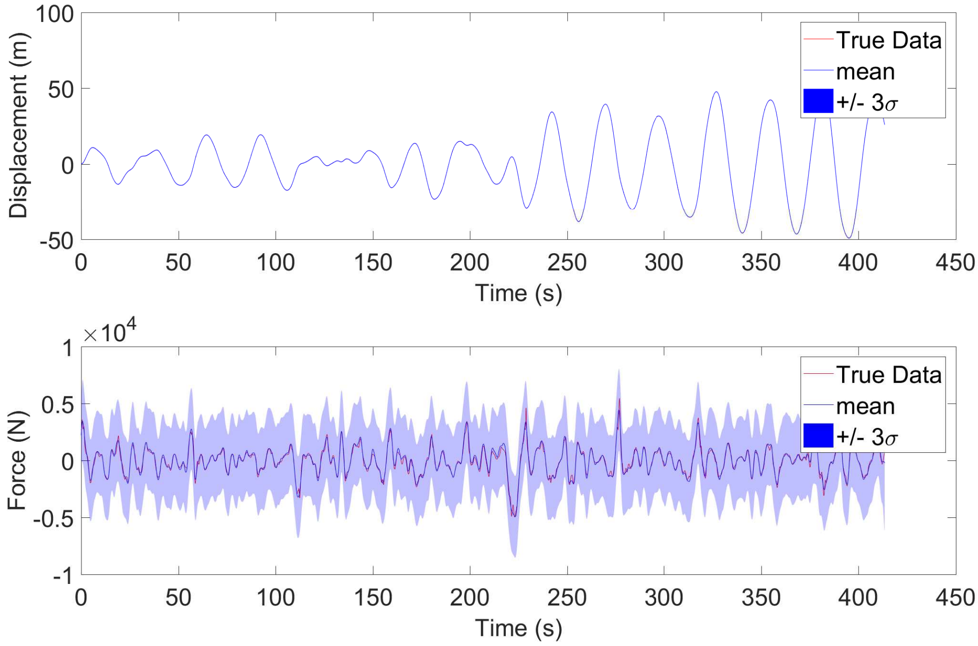

Figure 4: Smoother estimate of displacement and forcing signal from the LFM OMA on the system with true values shown in red, mean estimates as a solid blue line and three standard deviation confidence in the shaded regions

When running the model where only the mass is known, samples are generated from the posteriors over the GP hyperparameters and the system parameters, k and c. Plots of the histograms and CDFs of the parameters are shown in Fig. 3, and a summary of the results is shown in Table 2. It can be seen that the parameter estimates of the system parameters are within 2% of the ground truth for both the mean values of the distributions and the CDF estimates; this would indicate that the distributions are uni-modal and are not heavily skewed. When the values from the CDF estimates are used to run the filter-smoother over the full dataset, the predictions in Fig. 4 are made. The prediction over the force continues to be accurate with an NMSE on the mean of the signal of 3.1%. This result is comparable to the results obtained when the system properties are known — which is unexpected and encouraging. It follows that the estimates of the signal variance and length-scales are similar for this model and the model with known system parameters. It is also reassuring that the values for the GP hyperparameters are similar to those from the case where the system is known. This would suggest that the solution is stable and that the use of priors has alleviated some of the identifiability issues arising from the flexibility of the GP prior.

7 Conclusions

The work presented here has demonstrated through application to a simple system, a single-degree-of-freedom oscillator, the benefit of using a Gaussian Process Latent Force Model in the task of Operational Modal Analysis. Specifically, the state-space formulation of this model has allowed for efficient inference to be made, and as such, Markov-Chain Monte-Carlo methods can be used to recover distributions over the system parameters and the GP hyperparameters. It has been shown that these distributions can be used to elicit accurate point estimates of the parameters and also of the loading, something that is normally not possible in an OMA setting.

It is worth noting, however, that the tasks shown here have been performed on simulated data under a number of significant assumptions. First it is assumed that the linear model of the system can represent the behaviour well with low process noise. Secondly, the model is tested with a low level of artificial Gaussian white noise added. Thirdly, there exist priors over the system parameters in which there is non-negligible probability mass at the true values. Finally, it is assumed that the mass of the system is known accurately *a priori*; this is required to constrain the unidentifiability inherent in the model. The unidentifiability stems from the scaling of the forcing and the system parameters by the mass, that is, a system with the same response can be recovered if all the system parameters are multiplied by some constant α and if the forcing magnitude is also scaled by α. The flexibility of the GP as a prior over the forcing function only increases this issue. By fixing the mass and placing informative priors over the system parameters, the model can be constrained in such a way that it is more likely to converge to the true parameter posteriors.

Nevertheless, the authors believe that this represents a powerful *grey-box* approach to the problem of OMA. It is clear that the model is capable of returning good estimates of the forcing, and the system parameters under the constraints already discussed. The results presented here motivate further work into the use of GP LFM models in the modal analysis setting; this includes the extension of the current model to the multi-degree of freedom case. Additionally, in nonlinear systems where the structure of the nonlinearity is known (e.g. a cubic stiffness term), this model is also applicable although the Kalman filtering formulation must be replaced with a nonlinear counterpart, e.g. a particle filter. The last area of work is regarding the inference procedure; it is appealing to replace the MCMC inference used in this paper with a more efficient method which could be based on a state-augmentation approach or a two-step identification process where the system parameters and forcing are iterated recursively.

Acknowledgements

T. J. Rogers wishes to thank Ramboll Energy for their ongoing support of this research.

References

[1] Singiresu S Rao and Fook Fah Yap. *Mechanical Vibrations*, volume 4. Prentice Hall Upper Saddle River, 2011.

[2] Lennart Ljung. *System Identification*. Springer, 1998.

[3] Bart Peeters and Guido De Roeck. Stochastic system identification for operational modal analysis: a review. *Journal of Dynamic Systems, Measurement, and Control*, 123(4):659–667, 2001.

[4] Edwin Reynders. System identification methods for (operational) modal analysis: review and comparison. *Archives of Computational Methods in Engineering*, 19(1):51–124, 2012.

[5] Saeed Eftekhar Azam, Eleni Chatzi, and Costas Papadimitriou. A dual Kalman filter approach for state estimation via output-only acceleration measurements. *Mechanical Systems and Signal Processing*, 60:866–886, 2015.

[6] Saeed Eftekhar Azam, Eleni Chatzi, Costas Papadimitriou, and Andrew Smyth. Experimental validation of the Kalman-type filters for online and real-time state and input estimation. *Journal of Vibration and Control*, 23(15):2494–2519, 2017.

[7] Lennart Ljung. Asymptotic behavior of the extended Kalman filter as a parameter estimator for linear systems. *IEEE Transactions on Automatic Control*, 24(1):36–50, 1979.

[8] E Lourens, Costas Papadimitriou, Steven Gillijns, Edwin Reynders, Guido De Roeck, and Geert Lombaert. Joint input-response estimation for structural systems based on reduced-order models and vibration data from a limited number of sensors. *Mechanical Systems and Signal Processing*, 29:310–327, 2012.

[9] Mauricio Alvarez, David Luengo, and Neil Lawrence. Latent force models. In *Artificial Intelligence and Statistics*, pages 9–16, 2009.

[10] Jouni Hartikainen and Simo Sarkka. Sequential inference for latent force models. *arXiv preprint arXiv:1202.3730*, 2012.

[11] Anthony O'Hagan and John Frank Charles Kingman. Curve fitting and optimal design for prediction. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 1–42, 1978.

[12] Carl Edward Rasmussen and Christopher K. I. Williams. *Gaussian Processes for Machine Learning*. Citeseer, 2006.

[13] Rudolph Emil Kalman. A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, 82(1):35–45, 1960.

[14] Herbert E Rauch, CT Striebel, and F Tung. Maximum likelihood estimates of linear dynamic systems. *AIAA journal*, 3(8):1445–1450, 1965.

[15] Mauricio A Alvarez, David Luengo, and Neil D Lawrence. Linear latent force models using Gaussian processes. *IEEE transactions on pattern analysis and machine intelligence*, 35(11):2693–2705, 2013.

[16] Joaquin Quiñonero-Candela and Carl Edward Rasmussen. A unifying view of sparse approximate Gaussian process regression. *Journal of Machine Learning Research*, 6(Dec):1939–1959, 2005.

[17] Thang D Bui, Josiah Yan, and Richard E Turner. A unifying framework for sparse Gaussian process approximation using power expectation propagation. *arXiv preprint arXiv:1605.07066*, 2016.

[18] Mauricio Alvarez and Neil D Lawrence. Sparse convolved Gaussian processes for multi-output regression. In *Advances in neural information processing systems*, pages 57–64, 2009.

[19] Jouni Hartikainen and Simo Särkkä. Kalman filtering and smoothing solutions to temporal Gaussian process regression models. In *Machine Learning for Signal Processing (MLSP), 2010 IEEE International Workshop on*, pages 379–384. IEEE, 2010.

[20] Mark S Handcock and Michael L Stein. A Bayesian analysis of Kriging. *Technometrics*, 35(4):403–410, 1993.

[21] Michael L Stein. *Interpolation of Spatial Data: Some Theory for Kriging*. Springer Science & Business Media, 2012.

[22] Zoubin Ghahramani and Geoffrey E Hinton. Parameter estimation for linear dynamical systems. Technical report, Technical Report CRG-TR-96-2, University of Totronto, Dept. of Computer Science, 1996.

[23] Andrew Gelman, John B Carlin, Hal S Stern, David B Dunson, Aki Vehtari, and Donald B Rubin. *Bayesian Data Analysis*. CRC press, 2013.

[24] Simo Särkkä. *Bayesian Filtering and Smoothing*, volume 3. Cambridge University Press, 2013.

[25] G Najafian, RG Tickell, R Burrows, and JR Bishop. The UK Christchurch Bay compliant cylinder project: analysis and interpretation of Morison wave force and response data. *Applied Ocean Research*, 22(3):129–153, 2000.