Article:
Rogers, T.J. orcid.org/0000-0002-3433-3247, Worden, K. and Cross, E.J. (2020) On the application of Gaussian process latent force models for joint input-state-parameter estimation : With a view to Bayesian operational identification. Mechanical Systems and Signal Processing, 140. 106580. ISSN 0888-3270

https://doi.org/10.1016/j.ymssp.2019.106580

On The Application of Gaussian Process Latent Force Models for Joint Input-State-Parameter Estimation: With a View to Bayesian Operational Identification

T.J. Rogers, K. Worden, E.J. Cross

Dynamics Research Group, Department of Mechanical Engineering, University of Sheffield, Mappin Building, Mappin Street, Sheffield, S1 3JD

Please cite this version:

```
@article{rogers2020latent,
    title={On The Application of {Gaussian} Process Latent Force Models
           for Joint Input-State-Parameter Estimation: With a View to {Bayesian}
           Operational Identification},
    author={Rogers, T. J. and Worden, K. and Cross, E. J.},
    journal={Mechanical Systems and Signal Processing},
    year={2020},
    publisher={Elsevier},
    doi={https://doi.org/10.1016/j.ymssp.2019.106580}
}
```

On The Application of Gaussian Process Latent Force Models for Joint Input-State-Parameter Estimation: With a View to Bayesian Operational Identification

T.J. Rogers, K. Worden, E.J. Cross

Dynamics Research Group, Department of Mechanical Engineering, University of Sheffield, Mappin Building, Mappin Street, Sheffield, S1 3JD

Abstract

The problem of identifying dynamic structural systems is of key interest to modern engineering practice and is often a first step in an analysis chain, such as validation of computer models or structural health monitoring. While this topic has been well covered for tests conducted in a laboratory setting, identification of full-scale structures in place remains challenging. Additionally, during in service assessment, it is often not possible to measure the loading that a given structure is subjected to; this could be due to practical limitations or cost. Current solutions to this problem revolve around assumptions regarding the nature of the load a structure is subject to; almost exclusively this is assumed to be a white Gaussian noise. However, in many cases this assumption is insufficient and can lead to biased results in system identification. This current work presents a model which attempts the system identification task (in terms of the parametric estimation) in conjunction with estimation of the inputs to the system and the latent states — the displacements and velocities of the system. Within this paper, a Bayesian framework is presented for rigorous uncertainty quantification over both the system parameters and the unknown input signal. A Gaussian process latent force model allows a flexible Bayesian prior to be placed over the unknown forcing signal, which in conjunction with the state-space representation, allows *fully Bayesian* inference over the complete dynamic system and the unknown inputs.

Keywords: Bayesian, System Identification, Operational Modal Analysis, Gaussian Process, Latent Force Model

1. Introduction

Characterisation of dynamic behaviour of engineering structures and systems is a key step in a range of applications. Of particular importance is the use of these dynamic properties in structural health monitoring [1, 2], where modal testing can form the basis for many other methods, e.g. in civil infrastructure [3, 4]. Procedures for performing this testing within a controlled laboratory environment have been available for a number of years [5] (for linear systems). However, it is often impractical from a physical and/or cost perspective to test structures of interest in such a setting. Consider as an example an offshore platform; here the structure is unique, high value, and its loading is not easily quantified. In such a case it is impossible to perform laboratory-based experimental modal analysis, and controlling inputs to structures once installed can be challenging. Therefore, methods have been considered which attempt to perform the identification of the system after it is constructed and installed. This paper addresses this issue by building a Bayesian framework in which, from a number of output measurements, the parameters of a dynamic system and the unknown input can simultaneously be estimated.

Email address: tim.rogers@sheffield.ac.uk (T.J. Rogers)

Naturally, the problem of assessment of dynamic systems in operation has been approached previously in the literature; one current solution 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 [6, 7]; 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.

Likewise, the problem of input estimation has been previously considered in engineering. Lourens *et al.* [8] present a method based on Kalman filtering of a reduced-order model similar to the approach shown here but do not simultaneously perform parameter inference. Similarly Ma *et al.* [9] show results for estimation using a Kalman filter on a lumped-mass model of a beam with known parameters. Azam *et al.* [10] again apply a Bayesian filter-based system to model an output-only system without estimating the parameters of the model; this work is extended in [11]. Ching *et al.* [12] demonstrate the use of Bayesian filtering (comparing a particle filter and an extended Kalman filter) for parameter and state estimation when the inputs to the system are known. This problem is also considered in Gillijns and De Moor [13]. However, the task of joint input-state-parameter estimation remains challenging and is still an emerging area of investigation. In both Azam *et al.* [14] and Naets *et al.* [15] state-space approaches to the problem are considered with the use of state augmentation for parameter learning. More recently, Dertimanis *et al.* [16] make use of a combination of dual and augmented Kalman filters to address the joint input-state-parameter problem. Additionally, Maes *et al.* [17] develop an extended Kalman filter for the same task where the inputs and parameters of the system are augmented with the states to perform inference. Both of these papers demonstrate promising results and show the continued interest in this challenging and industrially relevant area of research.

In this work, as a solution to the problem of joint input-state-parameter estimation, the use of the Gaussian Process (GP) Latent Force Model (LFM) is proposed. The GPLFM was introduced by Alvarez *et al.* [18] as a tool for modelling data which are observations from a process whose outputs are governed by a second-order ordinary differential equation (ODE) which is forced by an unknown Gaussian process in time. The form of this model and its subsequent reformulation as a state-space model [19] is discussed in detail in Section 4. The authors of this paper have previously introduced the use of this model for joint input-state-parameter estimation for structural dynamical systems in [20], where the method is demonstrated on a single-degree-of-freedom system and the mass of the system is considered known. Independently of this, Nayek *et al.* [21] have recently shown the usefulness of this method in the case of input-state estimation only, i.e. the parameters of the system are considered known. The current work generalises the previous results of the authors to the case of multi-degree-of-freedom systems where all parameters are only known with weak prior knowledge. The advantage of this joint scheme is that it removes any bias in the forcing estimate from incorrect parameter values and vice versa.

This paper will review the necessary theory to develop the GPLFM, in Section 2 and Section 3. The model is described in full in Section 4 with its application to identification of structural systems considered in Section 5. The effectiveness of the proposed procedure is shown in Section 6 and these results are further discussed with conclusions made in Section 7. However, it is first helpful to provide an overview of the methodology and to establish notation which will be used throughout this paper.

1.1. An Overview of the Proposed Method

The method proposed here replaces the assumption of white noise forcing found in many methods with a more general case where the inputs to the system are considered unknown but that they can be modelled by a Gaussian process [22] in time, an approach first presented in Alvarez *et al.* [18].

By moving to a state-space representation of the system, it is possible to estimate efficiently the latent states of this model, which now include the displacements and velocities at each (modelled) degree of freedom in the structure and the unknown input forces. It transpires that this model can be computed as a Kalman filter/RTS smoother and then inference over the system can be performed in any usual manner.

Since the model is formed on the basis of placing a Bayesian prior over the input function, the full Bayesian inference solution is shown. The inference is achieved through a Markov Chain Monte Carlo inference over the parameters of the model which, as well as the structural parameters of the system, now include a small number of hyperparameters related to the forcing that the system has experienced. This quantifies the uncertainty in the system parameters and due to the form of the model allows a *distribution over functions*, which are possible realisations of the forcing time history, to be recovered.

The approach of the GP-LFM is shown to be a powerful method, which aims to identify system parameters and allows recovery of the forcing time history. The application of this identification represents a significant change in the way operational modal analysis is approached — increasing performance by modelling unknown inputs rather than assuming their form. In possession of posterior distributions over the parameters, further analysis can be conducted in a manner which rigorously quantifies and propagates uncertainty. In a similar manner, the distribution over the forcing can be propagated, e.g. into a fatigue damage accrual calculation.

1.2. Notation

A convention is adopted where a vector will be denoted by **bold** face, and matrices by CAPITALISATION. For sequential data, a subscript is used to indicate its time instant, t, for a sequence of variables the colon $(:)$ is used to indicate a range, i.e. for a sequence of length T, $\mathbf{x}_{1:T}$ denotes the variable $\mathbf{x}$ for the sequence from time points 1 to T. For matrix operations, $(\cdot)^{-1}$ indicates a matrix inverse and $(\cdot)^{\mathsf{T}}$ a matrix transpose. This work makes use of multivariate Gaussian densities, these are denoted as $\mathcal{N}(\boldsymbol{\mu}, \Sigma)$ with $\boldsymbol{\mu}$ being the mean vector and Σ being the covariance matrix. Dot notation is used as a shorthand for time derivatives throughout this paper, i.e. $\dot{\mathbf{x}}$ and $\ddot{\mathbf{x}}$ are the first and second derivatives of the vector $\mathbf{x}$ with respect to time. $\mathbb{I}$ is used to represent the identity matrix and $\mathbf{0}$ the zero matrix. When referencing the equations of a linear Gaussian state-space model, the usual convention of A being the state transition matrix, B being the control-input matrix, C being the state observation matrix, and D being the control-observation matrix is used. Additionally, Q is the covariance of the process noise and R the covariance of the observation noise.

2. Bayesian State Space Models

The Bayesian state-space model is a general framework for handling dynamic systems. The model is formulated on the representation of dynamic systems as sequences of latent (or hidden) states $\mathbf{x}$ which obey the Markov property. These states are related to a set of observed variables $\mathbf{y}$ which may be a subset of the hidden states — e.g. the displacement or velocity of an oscillator — or can be related to the hidden states through some probabilistic model — e.g. when observing the acceleration of an oscillator.

The Bayesian approach to solving this model is to assume that the functional relationships in the state-space model (SSM) are in fact stochastic and can be modelled as probability densities. The first distribution to consider is the transition density $f_\theta(\mathbf{x}_t \,|\, \mathbf{x}_{t-1}, \mathbf{u}_{t-1})$ which describes how the states evolve through time. The second fundamental part of the model specification is the observation density $g_\theta(\mathbf{y}_t \,|\, \mathbf{x}_t, \mathbf{u}_t)$ which models the relationship between the latent states $\mathbf{x}$ and the observed variables $\mathbf{y}$, when subjected to a set of control inputs $\mathbf{u}$. The subscript notation to

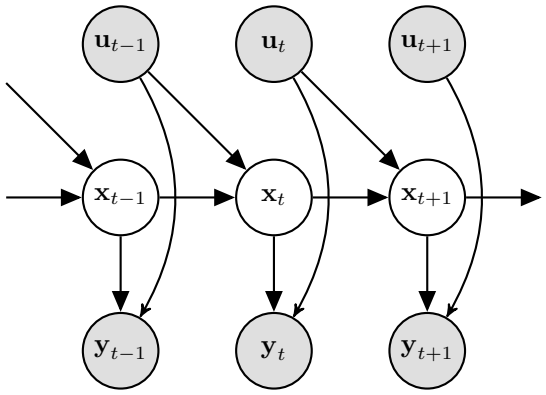


Figure 1: The graphical model of a general state-space model. The latent states $\mathbf{x}_{1:T}$ form a Markov process with the observation variables $\mathbf{y}_{1:T}$ conditioned only on the latent state at that time such that $p\left(\mathbf{y}_t \,|\, \mathbf{x}_{1:t}, \mathbf{u}_{1:t}\right) = p\left(\mathbf{y}_t \,|\, \mathbf{x}_t, \mathbf{u}_t\right)$.

these distributions indicates their dependence on some set of parameters — this is often omitted but important here since the determination of these parameters will be a part of the end goal. It can be show, trivially, how a specific function with a known noise model can be expressed as a probability density; the formulation of Bayesian linear regression is a simple example,

$$\mathbf{y} = X\mathbf{w}^\mathsf{T} + \boldsymbol{\varepsilon} \quad \boldsymbol{\varepsilon} \sim \mathcal{N}\left(0\,,\ \sigma_n^2 \mathbb{I}\right) \quad \Leftrightarrow \quad p\left(\mathbf{y} \,|\, X, \mathbf{w}, \sigma_n^2\right) = \mathcal{N}\left(X\mathbf{w}^\mathsf{T}\,,\ \sigma_n^2 \mathbb{I}\right) \tag{1}$$

The simple linear regression shown above on the left with a Gaussian noise can be equivalently written as the probability distribution over the output data $\mathbf{y}$ given the inputs X, the weights of the model $\mathbf{w}$, and the noise variance σ_n^2. The transition density and observation density fully describe an SSM if the initial conditions for the model are specified, a graphical model of a general SSM is shown in Fig. 1. This analysis gives rise to a general form of the SSM,

$$\mathbf{x}_t \sim f_\theta\left(\mathbf{x}_t \,|\, \mathbf{x}_{t-1}, \mathbf{u}_{t-1}\right) \tag{2a}$$
$$\mathbf{y}_t \sim g_\theta\left(\mathbf{y}_t \,|\, \mathbf{x}_t, \mathbf{u}_t\right) \tag{2b}$$

These equations form the basis of a very flexible model for time-series data, although SSMs can be applied to non-time-series data provided they represent a valid Markov process. The Bayesian state-space model is a popular tool and is widely adopted; a search of the Scopus database will return over 38,000 results with almost 35% of these being from engineering disciplines — good introductory texts on the subject can be found in [23–25].

When both the densities in Eq. (2) are Gaussian with means described by a linear function, the model is solvable in closed form. This solution is available for both the filtering density $p_\theta\left(\mathbf{x}_t \,|\, \mathbf{y}_{1:t}\right)$ and the smoothing density $p_\theta\left(\mathbf{x}_t \,|\, \mathbf{y}_{1:T}\right)$. These two densities are interesting to the modeller since they describe the distribution over the hidden states of the model at a point in time, given the sequence of observations up to the current time (in the case of the filtering density) or the total sequence of observations (in the case of the smoothing density).

It transpires that structural dynamical systems can be expressed in this form. That being the case, the general framework of Bayesian state-space models provides a strong theoretical foundation for solving structural dynamics problems. However, it is helpful that in the case of linear structural systems the state-space model is also (maybe unsurprisingly) linear. This being the case, and with assumptions of Gaussian noise, the solutions to this model are found in the popular Kalman

filtering and Rauch-Tung-Striebel (RTS) smoothing equations [26, 27]. These can be found in Appendix A.

These algorithms give rise to solutions for recovering the distributions over the hidden states of the model. However, these distributions remain dependent on the parameters of the model. These parameters are generally the entries in the A, B, C, D matrices (or *system matrices*); the noise covariances Q and R; and the initial conditions $\mathbf{x}_0$ and P_0. In the case of a structural dynamic system, the system matrices are defined in terms of the usual physical parameters of the system's mass, stiffness and damping; or canonically in terms of the natural frequencies and damping ratios of the system.

These parameters or subsets thereof can be collected into a set $\boldsymbol{\Theta}$ for notational convenience. The determination of these parameters remains as a challenge in modelling structures in operation. It may be the case that certain parameters in the model are known *a priori*; though, often many parameters remain unknown. Under the Bayesian paradigm, the learning of these parameters is not a task of finding a single set of *true* values but rather to infer a probability distribution over possible values. In general this distribution is referred to as the posterior distribution which is written down via Bayes' rule,

$$p\left(\boldsymbol{\Theta} \mid \mathcal{D}\right) = \frac{p\left(\mathcal{D} \mid \boldsymbol{\Theta}\right) p\left(\boldsymbol{\Theta}\right)}{p\left(\mathcal{D}\right)}$$

which expresses the posterior distribution as the ratio of the likelihood multiplied by the prior and normalised against the marginal likelihood or *model evidence*. With $\mathcal{D}$ being the observed dataset. The closed-form solutions expressed in the Kalman filter and RTS smoothing equations give access to such distributions over the latent states of the model, $\mathbf{x}_{1:T}$, which are unobserved given the observed quantities $\mathbf{y}_{1:T}$. In the nonlinear case, this task becomes intractable, necessitating the use of an approximate solution; however, this is beyond the scope of this work.

3. Gaussian Processes

The Gaussian process (GP) [22, 28] provides a Bayesian framework for nonparametric regression; this allows regression models of the form,

$$\mathbf{y} = f\left(X\right) + \boldsymbol{\varepsilon} \qquad \boldsymbol{\varepsilon} \sim \mathcal{N}\left(0\,,\, \sigma_n^2 \mathbb{I}\right)$$

to be learnt, where $\mathbf{y}$ is a vector of N observed targets, X a matrix of N observed inputs in D dimensions, and $\boldsymbol{\varepsilon}$ a vector of realisations from a zero-mean Gaussian white noise process with variance σ_n^2. A GP is fully defined by its mean and covariance function and represents a distribution over functions,

$$f(\mathbf{x}) \sim \mathcal{GP}\left(m\left(\mathbf{x}\right), k\left(\mathbf{x}, \mathbf{x}'\right)\right) \tag{3}$$

The mean function can be any parametric mapping from $\mathbf{x}$ to $f\left(\mathbf{x}\right)$, e.g. a polynomial. The correlation between the targets is captured by the covariance function which expresses the covariance between two input vectors $\mathbf{x}$ and $\mathbf{x}'$ through a Reproducing Kernel Hilbert Space (RKHS). In the GP, every observed target y_i for $i = 1, \ldots, \infty$ represents one dimension of a sample of a function from the GP. As such, any finite number of observations N of samples from the GP, $\mathbf{y}$, are observations from a multivariate Gaussian distribution with mean $m\left(X\right)$ and covariance $K\left(X, X\right)$ — which is the matrix of pairwise covariances calculated via the covariance function $k\left(\mathbf{x}, \mathbf{x}'\right)$.

To predict at a new test point $\mathbf{x}_\star$, or set of test points $X^\star$, the joint Gaussian distribution is extended to include the new outputs of the latent function $\mathbf{f}_\star$,

$$\begin{bmatrix}\mathbf{y}\\ f_\star\end{bmatrix} = \mathcal{N}\left(\begin{bmatrix}m\left(X\right)\\ m\left(\mathbf{x}_\star\right)\end{bmatrix},\ \begin{bmatrix}K\left(X,X\right)+\sigma_n^2\mathbb{I} & K\left(X,\mathbf{x}_\star\right)\\ K\left(\mathbf{x}_\star,X\right) & K\left(\mathbf{x}_\star,\mathbf{x}_\star\right)\end{bmatrix}\right) \tag{4}$$

Using standard results for multivariate Gaussian distributions, it is then possible to write down the distribution over $f_\star$, conditioned on the training data,

$$f_\star \sim \mathcal{N}\left(\mathbb{E}\left[f_\star\right],\ \mathbb{V}\left[f_\star\right]\right) \tag{5a}$$

$$\mathbb{E}\left[f_\star\right] = m\left(\mathbf{x}_\star\right) + K\left(\mathbf{x}_\star,X\right)\left(K\left(X,X\right)+\sigma_n^2\mathbb{I}\right)^{-1}\mathbf{y} \tag{5b}$$

$$\mathbb{V}\left[f_\star\right] = K\left(\mathbf{x}_\star,\mathbf{x}_\star\right) - K\left(\mathbf{x}_\star,X\right)\left(K\left(X,X\right)+\sigma_n^2\mathbb{I}\right)^{-1}K\left(X,\mathbf{x}_\star\right) \tag{5c}$$

If predicting at noisy output locations, i.e. $y_\star$, then it is trivial to add the noise variance σ_n^2 onto the predictive variance in Eq. (5c).

Clearly, the GP is dependent upon the kernel used. There are a number of choices available for the kernel function, each of which embeds a different prior belief as to which *family of functions* $\mathbf{f}$ is drawn from. For example, if a linear kernel is used the solution to a Bayesian linear regression is recovered. More commonly, nonlinear kernels will be chosen as many tasks require regression of nonlinear functions. Popular choices include the use of the Squared Exponential (SE) kernel or the Matérn class of kernels. These kernels, as well as being nonlinear, exhibit several other notable properties. They are stationary, meaning they do not depend on the absolute values of the two inputs, only the distance between them. Additionally, they are isotropic, i.e. they are insensitive to rotation, order, or translation of the inputs $\mathbf{x}$ and $\mathbf{x}'$. It is the authors' preference to use a Matérn 3/2 kernel, since Stein [29] argues that the SE kernel imposes unrealistic assumptions of smoothness when working with measured data. Heuristically, it has been the experience of the authors that this kernel is well suited to engineering data [30]. The form of this kernel will be shown later in the context of the Latent Force Model, where it has additional benefits. With regard to the mean function, it is usually set to zero in the literature, which is a reasonable assumption in many cases and one which is adopted here [22].

4. The Latent Force Approach

Alvarez *et al.* [18] first considered problems in which the data are assumed to be observations from a mechanistic system described by a second-order ordinary differential equation (ODE), the input to which is a Gaussian process, i.e.

$$M\ddot{\mathbf{z}} + C\dot{\mathbf{z}} + K\mathbf{z} = \mathbf{u}, \quad \mathbf{u}_j \sim \mathcal{GP}\left(0, k_j\left(t,t'\right)\right) \tag{6}$$

for $j = 1,\ldots,J$ degrees of freedom. This model was termed the Latent Force Model (LFM) and was applied to a number of datasets including pose estimation. However, in this form the model is computationally expensive and difficult to implement. A significant step was made by Hartikainen and Sarkka [19] in connecting their previous work [31] with the the LFM. In [31], it is shown how a zero-mean GP with a stationary kernel can be identically expressed as a linear Baysian state-space model which can be solved with the classic Kalman filtering [26] and Rauch-Tung-Striebel (RTS) smoothing [27] equations, these are shown in Appendix A.

Following Hartikainen and Särkkä [31] and converting the covariance function into an linear time invariant (LTI) stochastic differential equation requires the spectral density, by taking the Fourier

transform. First, however, it is necessary to present a linear time-invariant stochastic differential equation of order m which has the form shown in Eq. (7), where $f(t)$ is a random process of interest, $\mathbf{a}$ are some set of coefficients and $w(t)$ is a white noise process which has a spectral density $S_w(\omega) = q_c$.

$$\frac{\mathrm{d}^m f(t)}{\mathrm{d}t^m} + a_{m-1}\frac{\mathrm{d}^{m-1} f(t)}{\mathrm{d}t^{m-1}} + \ldots + a_1\frac{\mathrm{d}f(t)}{\mathrm{d}t} + a_0 f(t) = w(t). \tag{7}$$

This can be rearranged into a state-space model by constructing a vector $\mathbf{f(t)} = \left[f(t),\ \frac{\mathrm{d}f(t)}{\mathrm{d}t}, \ldots, \frac{\mathrm{d}^{m-1}f(t)}{\mathrm{d}t^{m-1}}\right]$ which gives rise to a first-order vector Markov process,

$$\frac{\mathrm{d}\mathbf{f(t)}}{\mathrm{d}t} = F_k\mathbf{f(t)} + L_k w(t), \tag{8}$$

where,

$$F_k = \begin{pmatrix} 0 & 1 & & \\ & \ddots & \ddots & \\ & & 0 & 1 \\ -a_0 & \ldots & -a_{m-2} & -a_{m-1} \end{pmatrix}, \qquad L_k = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}. \tag{9}$$

The coefficients $a_0, \ldots, a_{m-1}$ (Eq. (9)) are the coefficients of the denominator in the rational transfer function[1]. The rational transfer function of $\mathbf{f(t)}$ is given by,

$$H\left(i\omega\right) = \frac{b_0}{a_{m-1}\left(i\omega\right)^{m-1} + \ldots + a_1\left(i\omega\right) + a_0} \tag{10}$$

where the spectral density of the white noise process is given by the magnitude of the numerator squared, $q_c = |b_0|^2$. From this, the spectral density of the process can be written down as,

$$S(\omega) = H(i\omega)q_c H(i\omega)^{-1} \tag{11}$$

For an SSM in the form of Eq. (9), with an observation matrix $C_k = \begin{bmatrix}1 & 0 & \ldots & 0\end{bmatrix}^\mathsf{T}$, the spectral density of the Markov process can be written as,

$$S(\omega) = C_k\left(F_k + i\omega\mathbb{I}\right)^{-1} L_k q_c L_k^\mathsf{T}\left[\left(F_k + i\omega\mathbb{I}\right)^{-1}\right]^\mathsf{T} C_k^\mathsf{T} \tag{12}$$

It is also possible to work from a covariance function for a temporal process to calculate the spectral density; yielding a stable Markov process, provided the covariance is stationary.

For example, consider a GP defined with a Matérn covariance [22, 32] between any two points in time. Since this covariance is defined in terms of the absolute difference between the two input points, $r = |t - t'|$, it is stationary and, therefore, it is possible to perform a conversion to a Linear Gaussian State-Space Model (LGSSM).

[1] Although strictly the term transfer function is reserved for objects in the s-domain, here it is used to remain consistent with [19, 31] so as to avoid confusion. It is also helpful to keep in mind the distinction between the Frequency Response Function in Eq. (10), that of the Gaussian Process forcing model, and that of the structural system being identified.

The general Matérn covariance function is defined as,

$$k(r) = \sigma_f^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{\sqrt{2\nu}}{\ell} r \right)^{\nu} \mathcal{K}_\nu \left(\frac{\sqrt{2\nu}}{\ell} r \right). \tag{13}$$

where $\mathcal{K}_\nu$ is the modified Bessel function of the second kind. 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,

$$k(t, t') = \left(1 + \frac{\sqrt{3}r}{\ell} \right) \exp \left(-\frac{\sqrt{3}r}{\ell} \right). \tag{14}$$

For the general form of the Matérn covariance, the spectral density can be written down, by the Wiener-Khinchin theorem as [22],

$$S(\omega) = \sigma_f^2 \frac{2\pi^{\frac{1}{2}} \Gamma(\nu + 1/2)}{\Gamma(\nu)} \lambda^{2\nu} \left(\lambda^2 + \omega^2 \right)^{-(\nu - 1/2)}. \tag{15}$$

Setting $\lambda = \sqrt{2\nu} \Big/ \ell$ and $\nu = p + 1/2$, in the Matérn class of covariance functions, as before, it can be seen that,

$$S(\omega) \propto \left(\lambda^2 + \omega^2 \right)^{-(p+1)}. \tag{16}$$

This can be factorised simply, as shown in Eq. (17), to recover the transfer function needed to convert to a state-space model,

$$H(\mathrm{i}\omega) = (\lambda + \mathrm{i}\omega)^{-(p+1)}. \tag{17}$$

The spectral density of the white noise process $S_w(\omega) = q$ is equal to the constant in $S(\omega)$ (Eq. (15)) which is,

$$q = \frac{2\,\sigma_f^2\,\pi^{1/2} \lambda^{(2p+1)} \Gamma(p+1)}{\Gamma(p + 1/2)}. \tag{18}$$

For the case of the Matérn covariance with $\nu = 3/2$ as in Eq. (14), the equivalent state-space model is given by,

$$\frac{\mathrm{d}\mathbf{f(t)}}{\mathrm{d}t} = \begin{pmatrix} 0 & 1 \\ -\lambda^2 & -2\lambda \end{pmatrix} \mathbf{f(t)} + \begin{pmatrix} 0 \\ 1 \end{pmatrix} w(t). \tag{19}$$

It is then possible to convert this continuous-time state-space model to a discrete-time form and solve with a Kalman filter and RTS-smoother [23]. 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 [19]. This mechanistic model can be shown to include all SDOF and MDOF linear oscillators. The spectral density of the white noise process $w(t)$ in Eq. (19) is given by Eq. (18) substituting in $p = 1$.

Following the above procedure, it has been possible to show a state-space representation of the forcing on the system under the assumption that the forcing signal was generated by a zero-mean Gaussian process in time. To construct the LFM, Eq. (6) which defines the dynamical system is converted into its state-space form. In this standard procedure the second-order ODE that defines a mass-spring-damper system is converted into a continuous-time state-space form with $2R$ latent states corresponding to the displacements and velocities of each mass, i.e. $\mathbf{x} = \begin{bmatrix} z_1 & \dots & z_R & \dot{z}_1 & \dots & \dot{z}_R \end{bmatrix}^\intercal$. This system is written,

$$\dot{\mathbf{z}}(t) = \underbrace{\begin{bmatrix} \mathbf{0} & \mathbb{I} \\ M^{-1}K & M^{-1}C \end{bmatrix}}_{F_s} \mathbf{z}(t) + \underbrace{\begin{bmatrix} \mathbf{0} \\ M^{-1} \end{bmatrix}}_{G_s} \mathbf{u}(t) \tag{20}$$

The observation model can correspond to any sensor setup which is used, provided the system remains observable [33]. As such, the observed measurements $\mathbf{y}_t$, at time instant t, are given by,

$$\mathbf{y}_t = C\mathbf{z}_t + D\mathbf{u}_t + \mathbf{v}_t \quad \mathbf{v}_t \sim \mathcal{N}(0\,,\ R) \tag{21}$$

with C and D describing the sensor setup used and R defining the covariance of the measurement noise. Since the inputs to the system $\mathbf{u}(t)$ are considered unknown, the state vector $\mathbf{x}$ is augmented with additional states corresponding to these unknowns. The dimension of these augmented states is determined by the choice of covariance function used, which defines the state-space representation of the GP as shown above. In the case of the Matérn 3/2 kernel, which has been used as an example thus far, the number of augmented states is twice the number of forcing terms, e.g. the augmented states $\mathbf{x} = \begin{bmatrix} z_1 & \dots & z_R & \dot{z}_1 & \dots & \dot{z}_R & u_1 & \dot{u}1 \dots u_R & \dot{u}_R \end{bmatrix}$. Here $u_1, \dots, u_R$ are the forces entering at the R degrees of freedom in the model. $\dot{u}_1, \dots, \dot{u}_R$ correspond to the first time derivatives of those forces; this follows from the transfer function form of the Matérn kernel, i.e. if it were written as in Eq. (7). The added advantage of using a Matérn kernel at this point is that, unlike a squared exponential, it leads to a finite dimensional SSM representation, unlike a squared exponential, while retaining the flexibility to fit nonlinear regression models.

The full model of the system is now written down as,

$$\dot{\mathbf{x}}(t) = \underbrace{\begin{bmatrix} F_s & \tilde{G}_s \\ \mathbf{0} & F_k \otimes \mathbb{I}^{R\times R} \end{bmatrix}}_{F} \mathbf{x}(t) + L\mathbf{w}(t) \tag{22a}$$

$$\mathbf{y}_t = C\mathbf{x}_t + \mathbf{v}(t) \tag{22b}$$

where L is a column vector with zeros in every entry except those corresponding to $\dot{u}$ states where the entry is 1 and $\tilde{G}_s$ controls how the forces from the additional latent states enter into the dynamical system and $\otimes$ is the Kronecker product. Specifically,

$$\tilde{G}_s = \begin{bmatrix} \begin{bmatrix} 0 & 0 \\ m_1^{-1} & 0 \end{bmatrix} & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \begin{bmatrix} 0 & 0 \\ m_R^{-1} & 0 \end{bmatrix} \end{bmatrix} \tag{23}$$

As before, the model can be discretised by taking the matrix exponential of the continuous-time transition matrix multiplied by the time step, i.e. $A = \exp_m\{F.\Delta t\}$. The process noise of the

system is calculated from the spectral density of the continuous-time white noise process given by the kernel in Eq. (18). In discretising the model such that it can be solved with the Kalman filter and RTS smoother, it is necessary to calculate the covariance of the process noise in the discrete model, Q.

$$\begin{aligned} Q &= \int_{t_k}^{t_{k+1}} \exp_m\left(F.\left(t-\tau\right)\right) L_k q L_k^{\mathsf{T}} \exp_m\left(F.\left(t-\tau\right)\right)^{\mathsf{T}} \mathrm{d}\tau \\ &= P_\infty - A(\Delta t)\, P_\infty A(\Delta t)^{\mathsf{T}} \end{aligned} \tag{24}$$

where P_∞ is the steady-state solution to the SDE which is the solution to the Lyapunov equation [34],

$$F P_\infty + P_\infty F^{\mathsf{T}} + L Q L^{\mathsf{T}} = 0 \tag{25}$$

Now, in possession of a model which can represent the system, attention turns to the problem of inference. While methods such as the expectation maximisation algorithm exist for this form of model [35], these maximum likelihood solutions do not give information about the parameter uncertainties in the model and are not guaranteed to find the global maximum likelihood solutions. Instead, a fully Bayesian solution is adopted here 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 [36]. The parameter vector for this model, therefore, is $\boldsymbol{\theta} = \left[\mathbf{m}, \mathbf{k}, \mathbf{c}, \sigma_f^2, \ell, R\right]$. It is assumed that there is no process noise in the dynamics of the model, i.e. the model of the system as an MDOF oscillator accurately represents its behaviour. It should be noted that the use of a GP as the model of the latent force still introduces process noise to the system. Importantly, this process noise is a direct result of the formulation of the model and Eq. (24) shows how this can be calculated. This means that the matrix Q does not need to be estimated in full, rather it is completely defined by the dynamics of the system and spectral density q which in turn depends only upon the hyperparameters of the GP σ_f^2 and ℓ. The reader will notice that it has been assumed that there is no process noise in the dynamics of the system, i.e. that the MDOF model adopted is a correct representation of the system. For a physical structure this is a very strong assumption, therefore, the user has two options. First, additional process noise can be added to the dynamics of the system by estimating this directly through extending $\boldsymbol{\theta}$. If this is not done, then the estimated input force will be the inputs to the system minus the restoring force of the missing dynamics. However, if it is known that the discrepancy is small (e.g. the behaviour is well represented by the first M modes which are included in the model) then the estimation of the force will still be accurate and the inference procedure is simplified from a computational perspective.

The likelihood of the parameters in the model given the data observed for T time points, $p\left(\boldsymbol{\theta} \,|\, \mathbf{y}_{1:T}\right)$, is related to a quantity called the energy function $\varphi_T\left(\boldsymbol{\theta}\right)$ through a proportionality relationship — the energy function is the negative log likelihood of the filter scaled by a constant. If, rather than performing a fully Bayesian inference over the system the parameters were optimised it would be possible to use this to leverage the Bayesian Occam's Razor [37] via,

$$p\left(\boldsymbol{\theta} \,|\, \mathbf{y}_{1:T}\right) \propto \exp\left(-\varphi_T\left(\boldsymbol{\theta}\right)\right). \tag{26}$$

A user may decide that the need for the full posterior distributions is not significant but would rather obtain a "best" estimate of the parameters in the model. This could be the case if further analysis does not lend itself to uncertainty propagation. In this case the task becomes an optimisation problem; maximising the log marginal likelihood (equivalent to minimising the energy

function $\varphi_T(\boldsymbol{\theta})$) is a sensible choice of cost function for this task. The aim being to find the minimally complex model which explains the data. By initialising the energy function at the prior — $\varphi_0(\boldsymbol{\theta}) = p(\boldsymbol{\theta})$ — the value learnt in this optimisation will be the *maximum a posteriori* estimate of the parameters (subject to convergence of the optimisation routine).

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

$$\varphi_t(\boldsymbol{\theta}) = \varphi_{t-1}(\boldsymbol{\theta}) + \frac{1}{2}\log|2\pi S_t(\boldsymbol{\theta})| + \frac{1}{2}\mathbf{v}_t^{\mathsf{T}} S_t^{-1}\mathbf{v}_t. \tag{27}$$

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

$$\mathbf{v}_t = \mathbf{y}_t - C(A\mathbf{x}_{t-1}), \tag{28a}$$
$$S_t = C\left(AP_{t-1}A^{\mathsf{T}} + Q\right)C^{\mathsf{T}} + R, \tag{28b}$$

when P_{t-1} is the state covariance at $t-1$. The recursion for φ_T is started at $\varphi_0 = -\log p(\boldsymbol{\theta})$ [24]. Inference can then be performed using the standard MCMC Metropolis random walk algorithm [38], which for this problem is shown in Algorithm 1.

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

$n_b, n_s \leftarrow$ Length of burn in and desired number of samples
$\boldsymbol{\theta}_0 \leftarrow \left\{m_0, k_0, c_0, \sigma_{f0}^2, \ell_0\right\}$ ▷ Set Start Points for Markov Chain
$n_a \leftarrow 0, k \leftarrow 0$ ▷ Number of accepted $\boldsymbol{\theta}$, Number of Steps
$p\left(\boldsymbol{\theta}' \mid \boldsymbol{\theta}_k\right) = \mathcal{N}(0,\ \Sigma_p)$ ▷ Proposal is random walk with diagonal matrix Σ_p
$p(\boldsymbol{\theta} \mid \mathbf{y}) \propto \exp\{-\varphi_T(\boldsymbol{\theta})\}$ ▷ Posterior Likelihood is Proportional to the Energy Function
while $k < (n_b + n_s)$ **do**
  Sample $\boldsymbol{\theta}'$ from $p\left(\boldsymbol{\theta}' \mid \boldsymbol{\theta}_k\right)$
  Calculate $\varphi_T\left(\boldsymbol{\theta}'\right)$ ▷ Eq. (27)
  $\log(\alpha) = \min\left\{-\varphi_T\left(\boldsymbol{\theta}'\right) + \varphi_T(\boldsymbol{\theta}_k), 0\right\}$ ▷ Log Acceptance Ratio
  $\boldsymbol{\theta}_{k+1} \leftarrow \begin{cases} \boldsymbol{\theta}', & n_a{+}{+} \quad \text{with probability } \alpha \\ \boldsymbol{\theta}_{\mathbf{k}}, & \text{otherwise} \end{cases}$
  $k \leftarrow k + 1$
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 Modal Analysis

So far, a model has been shown for any generic second-order system, although this has been framed from the perspective of physical coordinates, this is equally applicable to any transformed coordinate system. The most popular choice in the literature is the move to *modal coordinates* — for example see [8, 10, 17]. This has been shown to be a very effective representation of the system which requires only the specification of the number of modes of interest. The inference in this case can be carried out in the same manner as before except that the parameters in the

model become the natural frequencies and damping parameters (modal damping ratios in the case of proporational damping or the modal damping matrix when this is not the case) and the latent forces learnt will be the so-called *modal forces*. These can be converted back to recover the physical forces through a transformaton using the mode shapes of the system. The normalised mode shapes must also be estimated in this case, assuming they are also unknown. If this procedure is adopted instead of that shown here the inference will change such that $\boldsymbol{\theta} = \left\{\Omega, Z, \Phi, \sigma_f^2, \ell, R\right\}$ where Ω is the diagonal matrix of eigenvalues (squared natural frequencies), Z is the modal damping matrix (which may or may not be diagonal), and Φ is the matrix of normalised mode shapes. Both of these possible settings will be discussed in the following subsections with reference to the implication this has on the measurement models, the observability of the system, and for specification of the priors.

5.1. The Measurement Model

The measurement model is key to the system under consideration. There are various choices available for this model which result in different specifications of the C matrix of the system, note that since no external forces are observed the B and D matrices are not included. This allows observation of the displacements, velocities, or accelerations of the system; or any combination thereof. In the case of a system expressed in modal coordinates, the C matrix will also need to include the mode shapes of the system, to relate the hidden states of the filter/smoother (the modal displacements and velocities) to these observed quantities. Importantly, for any of these combinations of measurements, the system remains a linear Gaussian state-space model.

For example, if observing acceleration measurements of the system, at a number of locations equal to the model order being considered, in a modal coordinate system, the C matrix for the model would be written,

$$C = \begin{bmatrix} \Phi\Omega^2 & \Phi Z & \Phi\Phi^{\mathsf{T}} & 0 \end{bmatrix} \tag{29}$$

considering that the internal states of the model are now,

$$\mathbf{x} = \begin{Bmatrix} \mathbf{q} \\ \dot{\mathbf{q}} \\ \mathbf{u} \\ \dot{\mathbf{u}} \end{Bmatrix} \tag{30}$$

where $\mathbf{q}$ and $\dot{\mathbf{q}}$ are the modal displacements and velocities respectively. $\mathbf{u}$ and $\dot{\mathbf{u}}$ remain the latent forces being inferred but are now the forces acting on the modal coordinate system rather than the physical one. Similar expressions can be developed for almost any set of measurements likely to be collected from a physical structure.

5.2. Observability and Identifiability

A central concern though should be, will the measurements actually be useful for recovering the system parameters and unknown inputs? The mathematical underpinning for the answer to this question is the observability of the filter itself. Since the states do not include the parameters being inferred and inference is performed over the whole collected time-series, the system remains linear. For linear Gaussian state-space models there are well established methods for checking the observability of the system being considered [23, 33]. As such, given the systems inferred it is possible to test for the observability of the system. The results of this are intuitive; if the components of the system observed are not influenced by the forces being inferred it will

be impossible to recover information about the system. While this addresses the recovery of the unknown input forces from the system, there remains a question of identifiability in the parameters.

The parameters of the model (in any coordinate system) will be inherently non-identifiable. This is due to the link between the forcing magnitude and the effective masses of the system. There are several ways in which this can be overcome: firstly one could specify prior distributions which ensure that the system is unique. Secondly one could observe in the posterior distributions the multi-modality of the solution when placing very weak priors. Thirdly, it is possible to assume certain parameters to be known. The third option is most commonly followed in literature, e.g. [14, 15] and in many cases is reasonable and the most practical solution. The second approach, although interesting poses a number of significant problems. The most important of which is that the inference procedure using MCMC becomes very difficult in terms of specifying the proposal covariance and ensuring convergence of the model. The first solution would be the one recommended by the authors of this paper, how to specify such prior distributions is discussed now.

5.3. Specification of Prior Distributions

A core part of any Bayesian inference is the specification of the prior distributions for the model. These express the user's belief about the system before observing any data. In the context of dynamic system identification, this information is not normally hard to come by, the strength of belief however should be considered. In a mechanical/civil engineering context it is rare that the system being identified will not be known at least to some extent. This information may exist in a very tangible form, for example drawings or models of the system from which parameters or the model order can be deduced as a starting point. There are also much weaker forms of prior information which can be exploited. An example of this would be that for certain structures where, by inspection of the output spectra, an engineer may be able to express belief about the sort of system; e.g. it is simple to translate *"I can see a peak that looks like a mode close to 0.5* Hz*"* into a formal prior distribution. Broader engineering knowledge can also be exploited; it is expected that systems will be significantly underdamped, for this reason it does not make sense to allow a prior distribution which would produces systems for which this is not true. Likewise, none of the parameters being inferred will be negative. For different systems the availability of this knowledge will be different, it is therefore, difficult to place formal rules on how the priors should be elicited. It can be said though that any user should ensure that the prior is specified in such a way that it truly represents their belief. In the absence of any knowledge regarding the system it may be useful to initialise this prior based upon some alternative analyses, i.e. adopt an empirical Bayes [39] approach based on some other identification algorithm, possibly one which excludes the input estimation to improve identifiability.

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 if available from testing — the authors have not found an instance of this in the literature but it is a simple intuitive jump.

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 uniform prior within bounds or an improper prior, for example $p\left(\sigma_f^2\right) = 1$. However, it can be helpful (in the authors' experience, at least) 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. This, again, is a natural leap from understanding the role of the signal variance in the specification of the Matérn kernel.

It is worth considering that the choice of kernel in the Gaussian process imposes additional prior belief upon the model. This should reflect the broad beliefs of the user regarding the properties of the function which may constitute the forcing; this includes the roughness of the function that may be expected. The limit of this is the choice of the squared-exponential kernel [22] which

is formally smooth, i.e. infinitely differentiable. The problem with this assumption is that it results in an infinite number of hidden states in the state-space representation. Fortunately, this is rarely a problem in practice, as a higher-order Matérn kernel tends quickly towards the squared-exponential and for physical data the smoothness assumption in the squared exponential is often unrealistic [29]. Often, since little is known about the forcing function the prior selection of the kernel, in practice, reduces to a limited choice. This choice imposes weak belief on the forcing and can be shown to be sufficiently expressive, especially if a "universal kernel" is chosen [40]. Common choices of kernel which may be considered could be, a white-noise kernel which imposes a white-noise prior on the loading, or a kernel from the Matérn family with an appropriate degree of smoothness controlled by the ν parameter [22].

This section has discussed how the users' prior belief regarding the dynamic system and the unknown forcing can be expressed formally in the construction of the model and through choice of prior probability distributions. As in all Bayesian inference, this is an important step. Invariably the choice of prior will affect the posterior distributions inferred and this should be considered. The authors would strongly urge against the use of improper priors in this type of model for two reasons; firstly it will lead to significant difficultly in inference and nonsensical results; secondly, it fails to capture the engineer's true belief since normally some reasonable assumptions about the system can be made, i.e. the system is underdamped. In the following section an example is presented where the priors are similar to those that could be reasonably elicited from available engineering knowledge about the system, an example of this could be design specifications or any available models.

6. Results

It remains to demonstrate the proposed methodology. To do so results are shown on a three-degree-of-freedom system with a forcing entering at mass one. The system is simulated with known "ground truth" parameters using a 5^{th}-order Runge-Kutta time step integration method. The system is loaded with a force signal measured from the Christchurch Bay Tower wave loading experiment [41] as a representative loading signal which is narrowband, with the sample rate of the system set to 10Hz. The length of the loading signal is 4135 time points, a second separate loading signal is used as an independent test of the input identification technique given the identified parameters. The system is defined as a three-degree-of-freedom lumped mass model with parameters shown in Table 1; the priors used in the inference procedure are also shown. The simulated outputs of the system are assumed to be the displacements at each mass which are corrupted with a Gaussian white-noise with a standard deviation of 1×10^{-3} mm.

The GP-LFM operates on the time series of the system, in this case, it is assumed that the displacements of all three masses are available, as discussed, by changing the measurement model, alternative measurements can be used. An additional example is shown in Appendix C where the system is partially observed, only the accelerations at masses two and three are considered available. This is of interest as the measurement set does not include measurements co-located with the inputs. The time series of the simulated system is shown in Fig. 2, it can be clearly seen that the forcing signal used is not a realisation of a Gaussian white-noise process — which is consistent with knowledge that wave-loading signals will correspond to a narrowband loading signal.

The power spectra of these signals reveals the non-Gaussian nature of the loading further and highlights the difficultly that this assumption would cause. It is seen that the spectrum of the loading is narrowband with a number of peaks below 0.5Hz. These peaks are also seen close to the first mode in the output spectra. In addition to this, the narrowband nature of the signal leads to a low signal-to-noise ratio in the other two modes, especially the third mode. This is more noticeable when considering the receptance FRF shown in Fig. 4, where the third mode appears

Parameter	Value	Unit	Prior
m_1	2000	t	$\log(m_1) \sim \mathcal{N}(\log(2000),\ 0.01)$
m_2	1800	t	$\log(m_2) \sim \mathcal{N}(\log(2000),\ 0.01)$
m_3	2200	t	$\log(m_3) \sim \mathcal{N}(\log(2000),\ 0.01)$
c_1	50	kN s	$\log(c_1) \sim \mathcal{N}(\log(30),\ 1)$
c_2	30	kN s	$\log(c_2) \sim \mathcal{N}(\log(30),\ 1)$
c_3	20	kN s	$\log(c_3) \sim \mathcal{N}(\log(30),\ 1)$
k_1	10500	kN m^{-1}	$\log(k_1) \sim \mathcal{N}(\log(10000),\ 1)$
k_2	10000	kN m^{-1}	$\log(k_2) \sim \mathcal{N}(\log(10000),\ 1)$
k_3	9500	kN m^{-1}	$\log(k_3) \sim \mathcal{N}(\log(10000),\ 1)$
σ_n^2	-		$\log\left(\sigma_n^2\right) \sim \mathcal{N}(\log(0.001),\ 1)$
σ_f^2	-		$\log\left(\sigma_f^2\right) \sim \mathcal{N}(\log(2000),\ 0.5)$
ℓ	-		$\log(\ell) \sim \mathcal{N}(0,\ 3)$

Table 1: Ground truth system parameters used in simulation and prior distributions used for MCMC.

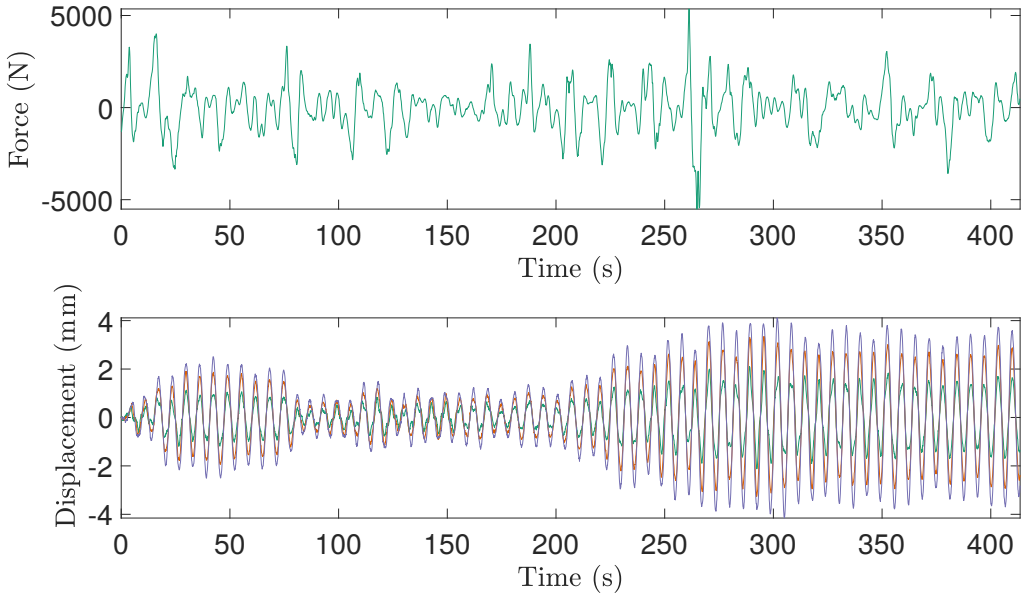


Figure 2: Figure showing the time-series data simulated from the system of interest with noise added. The upper plot shows the force signal used, the lower plot shows the measured displacements at masses one (green), two (orange), and three (purple).

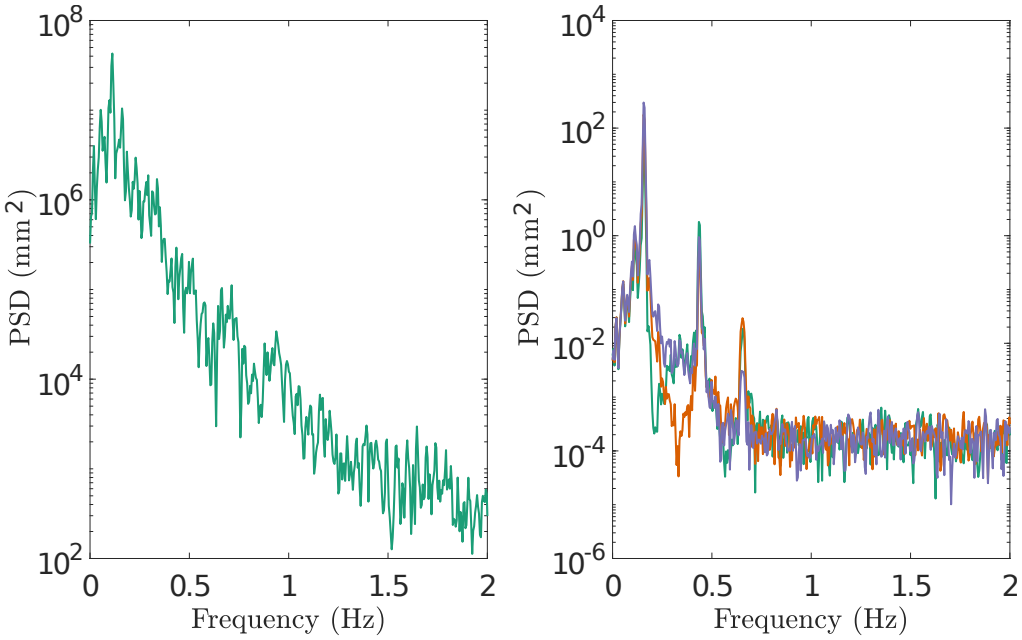


Figure 3: Figure showing the power spectra of the force (left-hand plot) and the measured displacements (right-hand plot). The spectra for masses masses one (green), two (orange), and three (purple) are shown.

with a significant amount of noise.

The non-Gaussian nature of the loading gives rise to artefacts in the output spectra which could be confused with modes of the structure and the narrowband content of the excitation signal leads to a poor signal-to-noise ratio in mode three of the model. The proposed method based on an MCMC inference scheme run over the GP-LFM is shown here to approach these problems.

It is assumed that the number of degrees of freedom of interest in the system (i.e. the expected number of poles and not necessarily physical degrees of freedom) is available *a priori*. Given that it is possible to perform MCMC inference over the full SSM, the fully Bayesian solution can be explored. The aim being to recover the parameter distributions of interest $\boldsymbol{\theta}$, as defined in Section 4. Here, this includes the distributions over the three masses, stiffnesses, and damping coefficients as well as the noise variance and GP kernel hyperparameters (signal variance and lengthscale) — equally this could be the natural frequencies, damping ratios and mode shapes if working in modal coordinates. It is assumed that the process noise in the system only arises from the GP used as the forcing function, and is therefore fully defined by the kernel hyperparameters. This assumption means that any discrepancy between the adopted model and the true dynamics will be modelled as part of the forcing signal being inferred.

MCMC inference is performed over the system using the priors specified, to ensure the parameters remain positive. This additive random walk in the log space with Gaussian priors is equivalent to a log-normal multiplicative random walk with log-normal priors in the original space. As is common, the MCMC random walk covariance is set based upon the acceptance rates of the chains and metrics such as the $\hat{R}$ [36]. The likelihood used is the exponential of the negative energy function of the Kalman filter run over the system [24]; however, for stability all of the likelihood calculations remain in log space for computing the acceptance ratio. This can be done in a "black-box" manner, i.e. it does not require any knowledge about the system and is based solely on its own metrics.

The results of this parameter estimation are shown in Fig. 5. Since there is no ground truth for the kernel hyperparameters, this is not marked, but the known ground truth values for the system

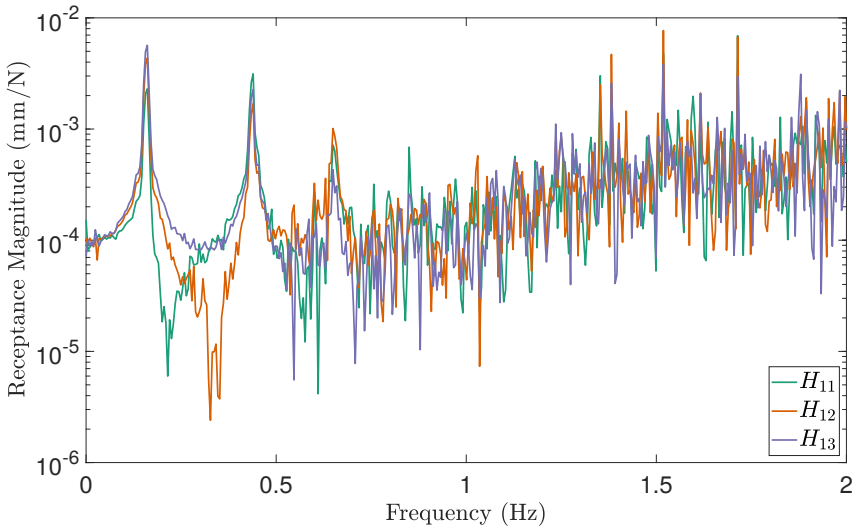
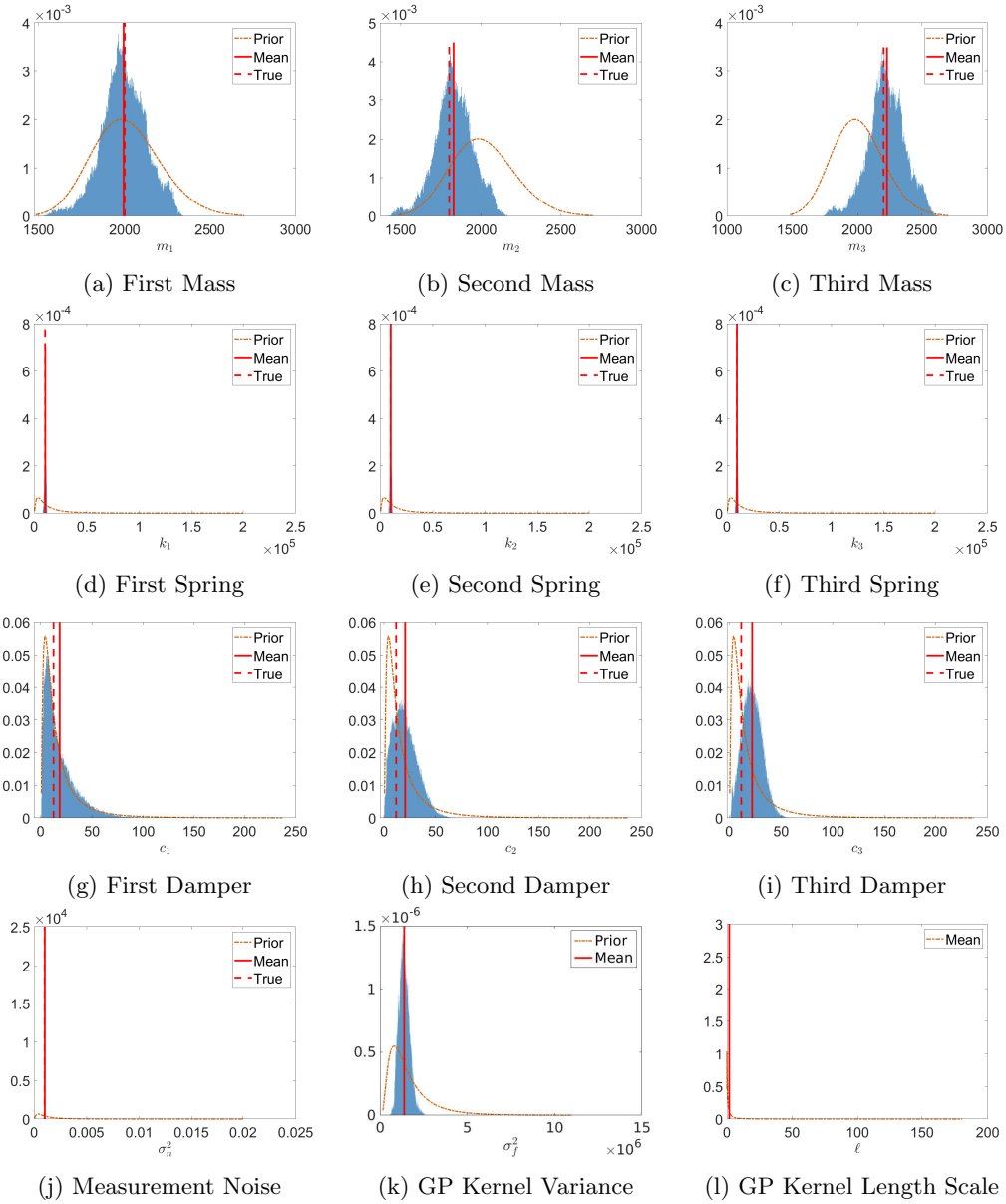


Figure 4: The receptance FRF of the system is shown for all three masses, force only enters at mass one.

parameters and measurement noise variance are shown along with the means for all distributions. It can be seen that all the ground truth values (where known) lie well within the probability mass of these distributions, all within 1σ. As such the parameters of the system can be recovered with rigorous quantification of their associated uncertainty.

The priors used in this inference have also been superimposed over the empirical distributions, up to a confidence of three standard deviations in all plots. At this point, it is worth at this point discussing the sensitivity of the various parameters to the priors and the interpretation of this from an engineering point of view. It is clear that the stiffness parameters in the system are well identified by the MCMC procedure. The means of the posterior distributions shift to be very close to the ground truth values and the variance of the posterior distributions is significantly reduced from the prior. Similar behaviour is seen in the posterior distributions over the mass parameters. The means of the distributions shift such that they agree very well with the known ground truth. A smaller reduction in uncertainty is seen in these than in the stiffness parameters but this is still encouraging that the inference is able to find reasonable posterior distributions. Turning attention to the viscous damping parameters, it is seen that there is a high degree of uncertainty in the model and higher discrepancy between the mean of the samples and the ground truth. This result is in line with the authors' experiences where damping is normally the most difficult quantity to identify. The measurement noise, similarly to the stiffness parameters is identified with both a very high degree of accuracy in the mean and with very small posterior variance. Considering the hyperparameters of the GP, σ_f^2 and ℓ, the length scale parameter posterior has very low posterior variance as may be expected due to the narrowband frequency content in the loading signal. The signal variance σ_f^2 sees a reduction of variance from the prior but the result has a far lower degree of confidence than the mass, for example. What is seen in these two sets of uncertainty is the inherent coupling between the mass and the magnitude of the forcing — σ_f^2 is the GP hyperparameter which controls the variance over the forcing. The link between these two variables is intuitive and goes some way to indicating when this framework can and will break down. If there is insufficient prior knowledge, then this nonidentifiability will result in highly multimodal posteriors, or more likely it will not be possible to design an MCMC scheme which will converge due to the complex posterior.

In addition to the parameter estimation the GPLFM also returns the smoothing distribution over the forcing signal. This distribution is a probabilistic prediction of the forcing time series that the structure was subjected to. Using only the mean of the smoother estimate of the state corresponding to the forcing, it is possible to calculate a normalised mean squared error for the



(a) First Mass (b) Second Mass (c) Third Mass

(d) First Spring (e) Second Spring (f) Third Spring

(g) First Damper (h) Second Damper (i) Third Damper

(j) Measurement Noise (k) GP Kernel Variance (l) GP Kernel Length Scale

Figure 5: Posterior distributions over system parameters and Gaussian process hyperparameters

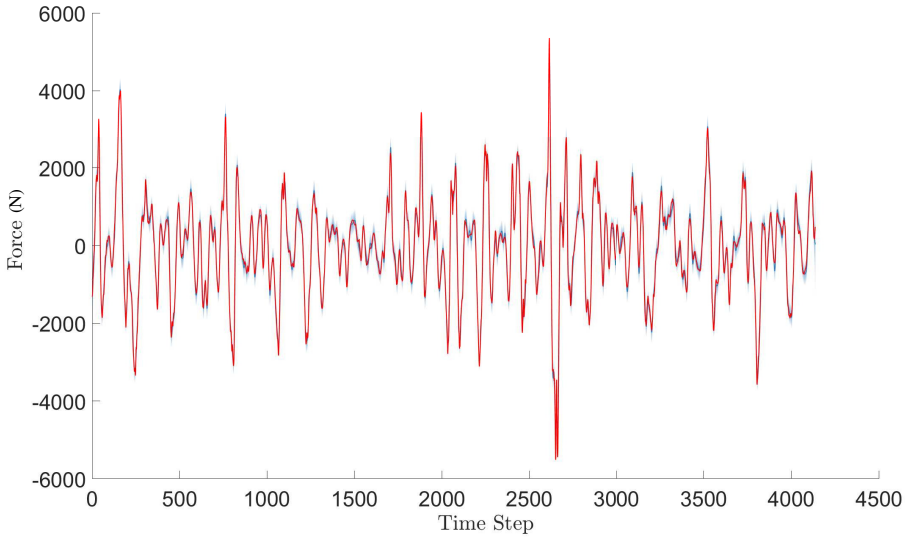


Figure 6: Estimation of the input signal on an independent test set

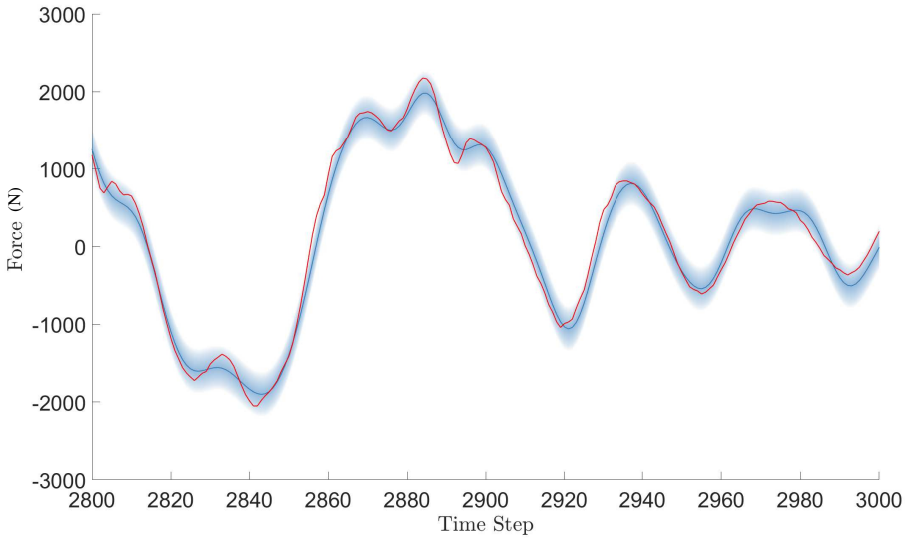


Figure 7: Small section of the estimation of the input signal on an independent test set

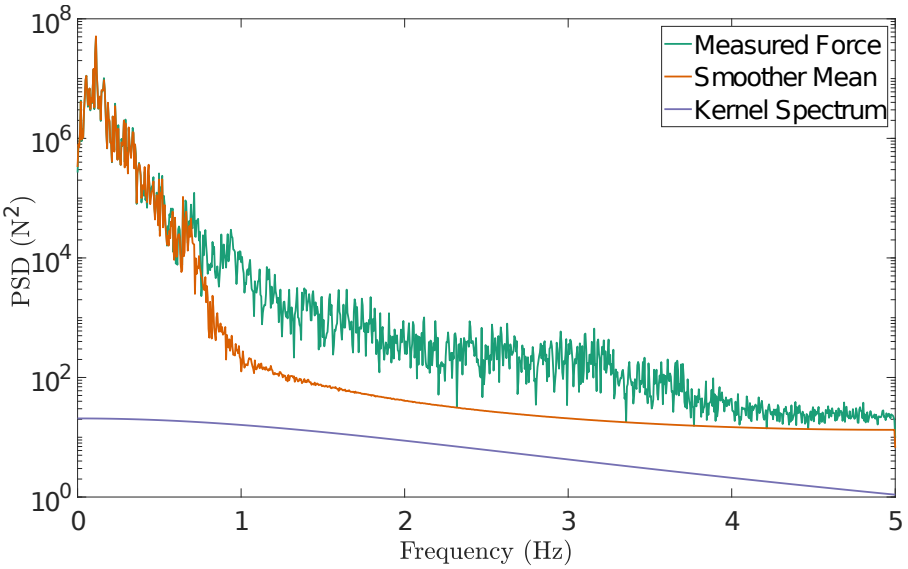


Figure 8: Power spectral densities of the measured force compared to the mean of the smoother estimate and the analytical spectrum of the prior.

estimate. This is done using the following formula,

$$NMSE = \frac{100}{N\sigma_y^2}\sum_{i=1}^{N}(\mathbf{y}_i - \hat{\mathbf{y}}_i)^2 \tag{31}$$

where $\mathbf{y}$ is the measured values of the outputs with variance σ_y^2, and $\hat{\mathbf{y}}$ are the estimated values for N data points. This score is easily interpretable where a value of zero corresponds to a *"perfect"* fit (in a deterministic sense) and a score of 100 is equivalent to taking the mean of the measured values as the prediction at every point. Using this metric, the estimation of the forcing signal from the data used for inferring the parameter distributions is 1.15. To further check the ability of the model to recover the forcing, the *maximum a posteriori* estimates of the parameters are used to infer the latent force in an unseen test set, i.e. when the same structural system is subjected to a different loading case also measured from the Christchurch Bay Tower. The predictions in this instance are shown in Figs. 6 and 7. For this new test case, the NMSE score is 1.22 which would indicate that the model is able to generalise well, given parameters inferred from only a short time-series input.

Considering the predictions seen in Figs. 6 and 7, it can be observed that the GPLFM has the effect of "smoothing" through the high-frequency components in the forcing signal. This is a consequence of the choice of covariance function upon which the model is formed. It has already been noted in Section 3 that the choice of covariance function encodes the prior belief as to the form of the function being modelled. The roughness of this function is one of the key characteristics of the covariance chosen. Here it is seen that the Matérn 3/2 covariance does not capture all the high-frequency components of the signal, instead these are modelled as the process noise. Since the model was derived via the spectral content of the covariance function it is possible to compare this to the spectra of both the measured force and the mean of the smoother estimate. This is done in Fig. 8, here it can be seen most clearly that the estimates of the forcing agree very well at low-frequency but that the high-frequency content of the forcing signal is not captured by the mean smoother estimate. In addition to this, the prior spectrum of the covariance function used can also be calculated, it is seen that this prior spectrum decays rapidly with increasing frequency owing to the fact that it contains the $\left(\lambda^2 + \omega^2\right)^{-2}$ term. The model, therefore, captures well the

frequency content of the forcing signal around the structural modes of the system but not at the higher frequency range. Capturing this content of the forcing would allow, if desired, samples of the estimated force to be used to perform traditional input-output analysis.

In this section, the ability of the GPLFM to recover through MCMC representative posterior distributions over the system parameters and also the times-series of the forcing has been shown. It has been shown that the system parameter distributions are reasonable considering the known ground truth values. Additionally, recovery of the forcing signal has a very low NSME score and the spectral content is seen to agree well at low frequency (around the structural modes) with the error coming from smoothing through the high-frequency content.

7. Conclusions and Future Directions

The work contained in this paper demonstrates that the GPLFM is a powerful model for performing output-only identification of structural dynamical systems. The model naturally handles cases in which the loading a system is subjected to is non-white-Gaussian — a situation in which traditional output-only methods struggle. The use of MCMC inference returns probabilistic posterior estimates over the system parameters and the input signal. The value of these distributions is increased as the results are propagated forward into further analyses, e.g. Bayesian finite element model updating. Here the assessment of outputs as a distribution can increase robustness and allow for risk-based decision making in any further analysis.

The use of models with a GPLFM structure additionally opens up further possibilities in the analysis of dynamical systems. It is possible to consider the impact of such a model on an experimental modal analysis, where it is known that there is significant measurement noise on the forcing signal. Additionally, the Bayesian framework may be exploited further to perform likelihood-based model selection. It is also worth considering methodologies for computation of such a model online in a Bayesian manner, since currently the estimation of the system parameters is performed in batch — although the input estimation can be performed in an online manner.

In conclusion, this paper has presented a complete model for joint input-state-parameter estimation under the Bayesian paradigm when the number of degrees of freedom in the model can be specified *a priori*. This has been shown to be robust to cases where the forcing on the system is a non-Gaussian signal which is not measured. The ability of the method to return posterior distributions over the system parameters, forcing input and measurement noise allows for propagation of uncertainty into future analyses. This represents a powerful approach for system identification and a step forward in output-only approaches in structural dynamics which opens up a number of further avenues of investigation.

Acknowledgements

Authors Rogers and Cross would like to gratefully acknowledge the support of the Engineering and Physical Sciences Research Council (EPSRC) for this work through grant number EP/S001565/1 and Worden through grant number EP/R003645/1.

References

[1] K. Worden, C. R. Farrar, G. Manson, and G. Park. The fundamental axioms of structural health monitoring. *Proceedings Of The Royal Society Of London A: Mathematical, Physical And Engineering Sciences* , 463(2082):1639–1664, 2007.

[2] C. R. Farrar and K. Worden. *Structural Health Monitoring: A Machine Learning Perspective*. John Wiley & Sons, 2012.

[3] B. Peeters and G. De Roeck. One-year monitoring of the Z24-Bridge: environmental effects versus damage events. *Earthquake Engineering & Structural Dynamics* , 30(2):149–171, 2001.

[4] J. M. Brownjohn, P.-Q. Xia, H. Hao, and Y. Xia. Civil structure condition assessment by FE model updating: methodology and case studies. *Finite Elements In Analysis And Design* , 37(10):761–775, 2001.

[5] D. J. Ewins. *Modal Testing: Theory and Practice*, volume 15. Research studies press Letchworth, 1984.

[6] B. Peeters and G. De Roeck. Stochastic system identification for operational modal analysis: a review. *Journal Of Dynamic Systems, Measurement, And Control* , 123(4):659–667, 2001.

[7] E. Reynders. System identification methods for (operational) modal analysis: review and comparison. *Archives Of Computational Methods In Engineering* , 19(1):51–124, 2012.

[8] E. Lourens, C. Papadimitriou, S. Gillijns, E. Reynders, G. De Roeck, and G. 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] C.-K. Ma, J.-M. Chang, and D.-C. Lin. Input forces estimation of beam structures by an inverse method. *Journal Of Sound And Vibration* , 259(2):387–407, 2003.

[10] S. E. Azam, E. Chatzi, and C. Papadimitriou. A dual Kalman filter approach for state estimation via output-only acceleration measurements. *Mechanical Systems And Signal Processing* , 60:866–886, 2015.

[11] S. E. Azam, E. Chatzi, C. Papadimitriou, and A. 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.

[12] J. Ching, J. L. Beck, and K. A. Porter. Bayesian state and parameter estimation of uncertain dynamical systems. *Probabilistic Engineering Mechanics* , 21(1):81–96, 2006.

[13] S. Gillijns and B. De Moor. Unbiased minimum-variance input and state estimation for linear discrete-time systems with direct feedthrough. *Automatica* , 43(5):934–937, 2007.

[14] S. E. Azam, V. Dertimanis, E. Chatzi, and P. C. Output-only schemes for joint input-state-parameter estimation of linear systems. In *Proceedings of the 1st ECCOMAS Thematic Conference on Uncertainty Quantification in Computational Sciences and Engineering*, pages 497–510, 2015.

[15] F. Naets, J. Croes, and W. Desmet. An online coupled state/input/parameter estimation approach for structural dynamics. *Computer Methods In Applied Mechanics And Engineering* , 283:1167–1188, 2015.

[16] V. K. Dertimanis, E. Chatzi, S. E. Azam, and C. Papadimitriou. Input-state-parameter estimation of structural systems from limited output information. *Mechanical Systems And Signal Processing* , 126:711–746, 2019.

[17] K. Maes, F. Karlsson, and G. Lombaert. Tracking of inputs, states and parameters of linear structural dynamic systems. *Mechanical Systems And Signal Processing* , 130:755–775, 2019.

[18] M. Alvarez, D. Luengo, and N. Lawrence. Latent force models. In *Artificial Intelligence and Statistics*, pages 9–16, 2009.

[19] J. Hartikainen and S. Sarkka. Sequential inference for latent force models. *ArXiv Preprint ArXiv:1202.3730* , 2012.

[20] T. Rogers, K. Worden, G. Manson, U. Tygesen, and E. Cross. A Bayesian filtering approach to operational modal analysis with recovery of forcing signals. In *Proceedings of ISMA 2018 - International Conference on Noise and Vibration Engineering and USD 2018 - International Conference on Uncertainty in Structural Dynamics*, 2018.

[21] R. Nayek, S. Chakraborty, and S. Narasimhan. A Gaussian process latent force model for joint input-state estimation in linear structural systems. *Mechanical Systems And Signal Processing* , 128:497–530, 2019.

[22] C. E. Rasmussen and C. K. I. Williams. *Gaussian Processes for Machine Learning*. The MIT Press, 2005.

[23] L. Ljung. *System Identification*. Springer, 1998.

[24] S. Särkkä. *Bayesian Filtering and Smoothing*. Cambridge University Press, 2013.

[25] D. Barber. *Bayesian Reasoning and Machine Learning*. Cambridge University Press, 2012.

[26] R. E. Kalman. A new approach to linear filtering and prediction problems. *Journal Of Basic Engineering* , 82(1):35–45, 1960.

[27] H. E. Rauch, C. Striebel, and F. Tung. Maximum likelihood estimates of linear dynamic systems. *AIAA Journal* , 3(8):1445–1450, 1965.

[28] A. O'Hagan and J. Kingman. Curve fitting and optimal design for prediction. *Journal Of The Royal Statistical Society. Series B (Methodological)* , pages 1–42, 1978.

[29] M. L. Stein. *Interpolation of Spatial Data: Some Theory for Kriging*. Springer, 1999.

[30] T. J. Rogers. *Towards Bayesian System Identification: With Application to SHM of Offshore Structures*. PhD thesis, University of Sheffield, 2018.

[31] J. Hartikainen and S. 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.

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

[33] M. S. Grewal. Kalman filtering. In *International Encyclopedia of Statistical Science*, pages 705–708. Springer, 2011.

[34] S. Särkkä and A. Solin. *Applied Stochastic Differential Equations*. Cambridge University Press, 2019.

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

[36] A. Gelman, J. B. Carlin, D. B. Rubin, A. Vehtari, D. B. Dunson, and H. S. Stern. *Bayesian Data Analysis, Third Edition*. CRC Press, 2013.

[37] C. E. Rasmussen and Z. Ghahramani. Occam's razor. In *Advances in neural information processing systems*, pages 294–300, 2001.

[38] S. Brooks, A. Gelman, G. Jones, and X.-L. Meng. *Handbook of Markov Chain Monte Carlo*. CRC Press, 2011.

[39] K. P. Murphy. *Machine Learning: A Probabilistic Perspective*. 2012.

[40] C. A. Micchelli, Y. Xu, and H. Zhang. Universal kernels. *Journal Of Machine Learning Research* , 7(Dec):2651–2667, 2006.

[41] G. Najafian, R. Tickell, R. Burrows, and J. 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.

Appendix A. Closed-Form Solutions to a Linear Gaussian SSM

Consider a linear Gaussian state-space model (LGSSM), which is defined by its transition and observation densities,

$$p\left(\mathbf{x}_t \,|\, \mathbf{x}_{t-1}, \mathbf{u}_{t-1}\right) = \mathcal{N}\left(A\mathbf{x}_{t-1} + B\mathbf{u}_{t-1}\,,\; Q\right) \tag{A.1a}$$

$$p\left(\mathbf{y}_t \,|\, \mathbf{x_t}, \mathbf{u}_t\right) = \mathcal{N}\left(C\mathbf{x}_t + D\mathbf{u}_t\,,\; R\right) \tag{A.1b}$$

The model is defined recursively from a prior distribution over the hidden states, $\mathbf{x}_0 \sim \mathcal{N}\left(\mathbf{m}_0\,,\; P_0\right)$, for T discrete time steps. To recover the predition step of the filter, the joint distribution between the hidden states at time t and $t-1$ is considered conditioned on the previous observations $\mathbf{y}_{1:t-1}$ and the control input $\mathbf{u}_{t-1}$. This joint distribution is given by,

$$p\left(\mathbf{x}_t, \mathbf{x}_{t-1} \,|\, \mathbf{u}_{t-1}, \mathbf{y}_{1:t-1}\right) = p\left(\mathbf{x}_t \,|\, \mathbf{x}_{t-1}, \mathbf{u}_{t-1}\right) p\left(\mathbf{x}_{t-1} \,|\, \mathbf{y}_{1:t-1}\right) \tag{A.2}$$

where the first density on the right hand side is simply the transition density in Eq. (A.1a) and the second density is the filtering density from the previous time step. Therefore, using the equations in Appendix B, this joint can be rewritten as,

$$p\left(\mathbf{x}_t, \mathbf{x}_{t-1} \,|\, \mathbf{u}_{t-1}, \mathbf{y}_{1:t-1}\right) = \mathcal{N}\left(\mathbf{x}_t | A\mathbf{x}_{t-1} + B\mathbf{u}_{t-1}\,,\; Q\right) \mathcal{N}\left(\mathbf{x}_{t-1} | \mathbf{m}_{t-1}\,,\; P_{t-1}\right) \tag{A.3a}$$

$$= \mathcal{N}\left(\begin{bmatrix} \mathbf{x}_{t-1} \\ \mathbf{x}_t \end{bmatrix} \,\middle|\, \begin{bmatrix} \mathbf{m}_{t-1} \\ A\mathbf{m}_{t-1} + B\mathbf{u}_{t-1} \end{bmatrix} \,,\, \begin{bmatrix} P_{t-1} & P_{t-1}A^{\mathsf{T}} \\ AP_{t-1} & AP_{t-1}A^{\mathsf{T}} + Q \end{bmatrix} \right) \tag{A.3b}$$

From Eq. (B.2), by marginalising out $\mathbf{x}_{t-1}$ the prediction density is,

$$\begin{aligned} p\left(\mathbf{x}_t \,|\, \mathbf{y}_{1:t-1}, \mathbf{u}_{t-1}\right) &= \mathcal{N}\left(\mathbf{x}_t | A P_{t-1}\,,\; AP_{t-1}A^{\mathsf{T}} + Q\right) \\ &= \mathcal{N}\left(\mathbf{x}_t | \bar{\mathbf{m}}_t\,,\; \bar{P}_t\right) \end{aligned} \tag{A.4}$$

$\bar{\mathbf{m}}_t$ and $\bar{P}_t$ are the predictive mean and covariance of the hidden states.

Attention turns to the measurement update, conditioning the hidden states at the current time t on the measurement collected at this time. Again this is achieved using simple Gaussian identities from Appendix B. First the joint distribution between $\mathbf{x}_t$ and $\mathbf{y}_t$ is formed,

$$p\left(\mathbf{x}_t, \mathbf{y}_t \,|\, \mathbf{u}_t, \mathbf{y}_{1:t-1}\right) = p\left(\mathbf{y}_t \,|\, \mathbf{x}_t, \mathbf{u}_t\right) p\left(\mathbf{x}_t \,|\, \mathbf{y}_{1:t-1}\right) \tag{A.5a}$$

$$= \mathcal{N}\left(\mathbf{y}_t | C\mathbf{x}_t + D\mathbf{u}_t\,,\; R\right) \mathcal{N}\left(\mathbf{x}_t | \bar{\mathbf{m}}_t\,,\; \bar{P}_t\right) \tag{A.5b}$$

$$= \mathcal{N}\left(\begin{bmatrix} \mathbf{x}_t \\ \mathbf{y}_t \end{bmatrix} \,\middle|\, \begin{bmatrix} \bar{\mathbf{m}}_t \\ C\bar{\mathbf{m}}_t + D\mathbf{u}_t \end{bmatrix} \,,\, \begin{bmatrix} \bar{P}_t & \bar{P}_t C^{\mathsf{T}} \\ C\bar{P}_t & C\bar{P}_t C^{\mathsf{T}} + R \end{bmatrix} \right) \tag{A.5c}$$

It can be seen that the derivation of this step follows closely that of the prediction step, forming the joint distribution between the observation density and the predictive density. Here, however, it is

necessary to condition $\mathbf{x}_t$ on $\mathbf{y}_t$ rather than marginalising out. This is simply another application of a Gaussian identity, here the conditioning identity Eq. (B.3). Therefore, the filtering density can be written down,

$$\begin{aligned} p\left(\mathbf{x}_t \,|\, \mathbf{y}_t, \mathbf{y}_{1:t-1}\right) &= p\left(\mathbf{x}_t \,|\, \mathbf{y}_{1:t}\right) \\ &= \mathcal{N}\left(\mathbf{x}_t | \mathbf{m}_t \,,\, P_t\right) \end{aligned} \tag{A.6}$$

Defining the Kalman gain $K = \bar{P}_t C^{\mathsf{T}} \left(C \bar{P}_t C^{\mathsf{T}} + R\right)^{-1}$, it is possible to write down,

$$\mathbf{m}_t = \bar{\mathbf{m}}_t + K\left[\mathbf{y}_t - C\bar{\mathbf{x}}_t\right] \tag{A.7}$$

$$P_t = \left(\mathbb{I} - KC\right)\bar{P}_t \tag{A.8}$$

therefore, through sequentially applying these prediction and update steps it is possible to recover the filtering density at time t, $p\left(\mathbf{x}_{1:t} \,|\, \mathbf{y}_{1:t}, \mathbf{u}_{1:t}\right)$.

The RTS smoother equations are defined in a similar manner, to recover the smoothing density of the model $p\left(\mathbf{x}_{1:T} \,|\, \mathbf{y}_{1:T}, \mathbf{u}_{1:T}\right)$. The smoother is formed by a backwards recursion from time step T to time zero, this recursion is started by setting the smoothing density at time T to be equal to the filtering density at time T. The joint distribution of $\mathbf{x}_t$ and $\mathbf{x}_{t+1}$ is formed,

$$\begin{aligned} p\left(\mathbf{x}_t, \mathbf{x}_{t+1} \,|\, \mathbf{y}_{1:t}\right) &= p\left(\mathbf{x}_{t+1} \,|\, \mathbf{x}_t\right) p\left(\mathbf{x}_t \,|\, \mathbf{y}_{1:t}\right) \\ &= \mathcal{N}\left(A\mathbf{x}_{t-1} + B\mathbf{u}_{t-1} \,,\, Q\right) \mathcal{N}\left(\mathbf{x}_t | \mathbf{m}_t \,,\, P_t\right) \\ &= \mathcal{N}\left(\begin{bmatrix} \mathbf{x}_t \\ \mathbf{x}_{t+1} \end{bmatrix} \,\middle|\, \begin{bmatrix} \mathbf{m}_t \\ A\mathbf{m}_t + B\mathbf{u}_t \end{bmatrix} \,,\, \begin{bmatrix} P_t & P_t A^{\mathsf{T}} \\ AP_t & AP_tA^{\mathsf{T}} + Q \end{bmatrix} \right) \end{aligned} \tag{A.9}$$

Due to the Markov property, $p\left(\mathbf{x}_t \,|\, \mathbf{x}_{t+1}, \mathbf{y}_{1:T}, \mathbf{u}_{1:T}\right) = p\left(\mathbf{x}_t \,|\, \mathbf{x}_{t+1}, \mathbf{y}_{1:t}, \mathbf{u}_t\right)$, this allows the distribution $p\left(\mathbf{x}_t \,|\, \mathbf{x}_{t+1}, \mathbf{y}_{1:T}, \mathbf{u}_{1:T}\right)$ to be written down through conditioning Eq. (A.9), as,

$$\begin{aligned} p\left(\mathbf{x}_t \,|\, \mathbf{x}_{t+1}, \mathbf{y}_{1:T}, \mathbf{u}_{1:T}\right) &= p\left(\mathbf{x}_t \,|\, \mathbf{x}_{t+1}, \mathbf{y}_{1:t}, \mathbf{u}_t\right) \\ &= \mathcal{N}\left(\mathbf{x}_t | \tilde{\mathbf{m}} \,,\, \tilde{P}\right) \end{aligned} \tag{A.10}$$

with,

$$G_t = P_t A^{\mathsf{T}} \left(AP_tA + Q\right)^{-1} \tag{A.11}$$

$$\tilde{\mathbf{m}} = \mathbf{m}_t + G_t\left(\mathbf{x}_{t+1} - A\mathbf{m}_t - B\mathbf{u}_t\right) \tag{A.12}$$

$$\tilde{P} = P_t - G_t\left(AP_tA^{\mathsf{T}} + Q\right)G_t^{\mathsf{T}} \tag{A.13}$$

Considering again the joint distribution between $\mathbf{x}_t$ and $\mathbf{x}_{t+1}$ but conditioned on all the observation data, the joint Gaussian density can be written down as,

$$p\left(\mathbf{x}_t, \mathbf{x}_{t+1} \,|\, \mathbf{y}_{1:T}, \mathbf{u}_{1:T}\right) = p\left(\mathbf{x}_t \,|\, \mathbf{x}_{t+1}, \mathbf{y}_{1:T}, \mathbf{u}_{1:T}\right) p\left(\mathbf{x}_{t+1} \,|\, \mathbf{y}_{1:T}, \mathbf{u}_{1:T}\right) \tag{A.14}$$

It can be seen that this joint density is the product of the density calculated in Eq. (A.10) and the smoothing density at time $t+1$. Denoting the smoother mean and covariance at time t as $\mathbf{m}_t^s$ and P_t^s, it is possible to write down this joint distribution,

$$p\left(\mathbf{x}_t, \mathbf{x}_{t+1} \,|\, \mathbf{y}_{1:T}, \mathbf{u}_{1:T}\right) = \mathcal{N}\left(\mathbf{x}_t | \tilde{\mathbf{m}}\,,\ \tilde{P}\right) \mathcal{N}\left(\mathbf{x}_{t+1} | \mathbf{m}_{t+1}^s\,,\ P_{t+1}^s\right) \tag{A.15}$$

$$= \mathcal{N}\left(\begin{bmatrix} \mathbf{x}_{t+1} \\ \mathbf{x}_t \end{bmatrix} \,\middle|\, \begin{bmatrix} \mathbf{m}_{t+1}^s \\ \tilde{\mathbf{m}} \end{bmatrix}\ ,\ \begin{bmatrix} P_{t+1}^s & P_{t+1}^s G_t^\mathsf{T} \\ G_t P_{t+1}^s & G_t P_{t+1}^s G_t^\mathsf{T} + \tilde{P} \end{bmatrix} \right) \tag{A.16}$$

Finally, applying the marginalisation identity it is possible to write down the smoothing distribution over $\mathbf{x}_t$,

$$p\left(\mathbf{x} \,|\, \mathbf{y}_{1:T}, \mathbf{u}_{1:T}\right) = \mathcal{N}\left(\mathbf{x}_t | \mathbf{m}_t^s\,,\ P_t^s\right) \tag{A.17}$$

where,

$$\mathbf{m}_t^s = \mathbf{m}_t + G_t\left(\mathbf{m}_{t+1}^s - A\mathbf{m}_t - B\mathbf{u}_t\right) \tag{A.18}$$

$$P_t^s = P_t + G_t\left(P_{t+1}^s - AP_tA^\mathsf{T} - Q\right) G_t^\mathsf{T} \tag{A.19}$$

Appendix B. Properties of Multivariate Gaussian Distributions

For two sets of random variables $\mathbf{a} \in \mathbb{R}^{N_a}$ and $\mathbf{b} \in \mathbb{R}^{N_b}$, their joint Gaussian distribution is considered.

$$\begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \boldsymbol{\mu}_a \\ \boldsymbol{\mu}_b \end{bmatrix}\ ,\ \begin{bmatrix} \Sigma_{aa} & \Sigma_{ba} \\ \Sigma_{ab} & \Sigma_{bb} \end{bmatrix} \right) \tag{B.1}$$

The marginal distributions are given by,

$$p\left(\mathbf{a}\right) = \mathcal{N}\left(\boldsymbol{\mu}_a\,,\ \Sigma_{aa}\right) \tag{B.2a}$$

$$p\left(\mathbf{b}\right) = \mathcal{N}\left(\boldsymbol{\mu}_b\,,\ \Sigma_{bb}\right) \tag{B.2b}$$

and the cross-covariances $\Sigma_{ab} = \Sigma_{ba}^\mathsf{T}$. It is also possible to write down the conditional distributions,

$$p\left(\mathbf{a} \,|\, \mathbf{b}\right) = \mathcal{N}\left(\boldsymbol{\mu}_a + \Sigma_{ab}\Sigma_{bb}^{-1}\left(\mathbf{b} - \boldsymbol{\mu}_b\right)\,,\ \Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba}\right) \tag{B.3a}$$

$$p\left(\mathbf{b} \,|\, \mathbf{a}\right) = \mathcal{N}\left(\boldsymbol{\mu}_b + \Sigma_{ba}\Sigma_{aa}^{-1}\left(\mathbf{a} - \boldsymbol{\mu}_a\right)\,,\ \Sigma_{bb} - \Sigma_{ba}\Sigma_{aa}^{-1}\Sigma_{ba}\right) \tag{B.3b}$$

For the product of two Gaussians, the result is also a Gaussian distribution with a known normalising constant which can be calculated in closed form,

$$\begin{aligned} p\left(\mathbf{x} \,|\, \boldsymbol{\mu}_a, \Sigma_{aa}\right) p\left(\mathbf{x} \,|\, \boldsymbol{\mu}_b, \Sigma_{bb}\right) &= \mathcal{N}\left(\boldsymbol{\mu}_a\,,\ \Sigma_{aa}\right) \mathcal{N}\left(\boldsymbol{\mu}_b\,,\ \Sigma_{bb}\right) \\ &= Z^{-1} \mathcal{N}\left(\left(\Sigma_{aa}^{-1} + \Sigma_{bb}^{-1}\right)^{-1}\left(\Sigma_{aa}^{-1}\boldsymbol{\mu}_\mathbf{a} + \Sigma_{bb}^{-1}\boldsymbol{\mu}_\mathbf{b}\right)\,,\ \left(\Sigma_{aa}^{-1} + \Sigma_{bb}^{-1}\right)^{-1}\right) \end{aligned} \tag{B.4}$$

with,

$$Z^{-1} = (2\pi)^{-D/2} \left|\Sigma_{aa} + \Sigma_{bb}\right|^{-1/2} \exp\left\{-\frac{1}{2}\left(\boldsymbol{\mu}_{\mathbf{a}} - \boldsymbol{\mu}_{\mathbf{b}}\right)^{\mathsf{T}} \left(\Sigma_{aa} + \Sigma_{bb}\right)^{-1} \left(\boldsymbol{\mu}_{\mathbf{a}} - \boldsymbol{\mu}_{\mathbf{b}}\right)\right\} \tag{B.5}$$

as such, the integral over the product with respect to $\mathbf{x}$ is given by,

$$\begin{aligned} &\int p\left(\mathbf{x} \,|\, \boldsymbol{\mu}_a, \Sigma_{aa}\right) p\left(\mathbf{x} \,|\, \boldsymbol{\mu}_b, \Sigma_{bb}\right) \mathrm{d}\mathbf{x} = Z^{-1} \\ &\quad = (2\pi)^{-D/2} \left|\Sigma_{aa} + \Sigma_{bb}\right|^{-1/2} \exp\left\{-\frac{1}{2}\left(\boldsymbol{\mu}_{\mathbf{a}} - \boldsymbol{\mu}_{\mathbf{b}}\right)^{\mathsf{T}} \left(\Sigma_{aa} + \Sigma_{bb}\right)^{-1} \left(\boldsymbol{\mu}_{\mathbf{a}} - \boldsymbol{\mu}_{\mathbf{b}}\right)\right\} \end{aligned} \tag{B.6}$$

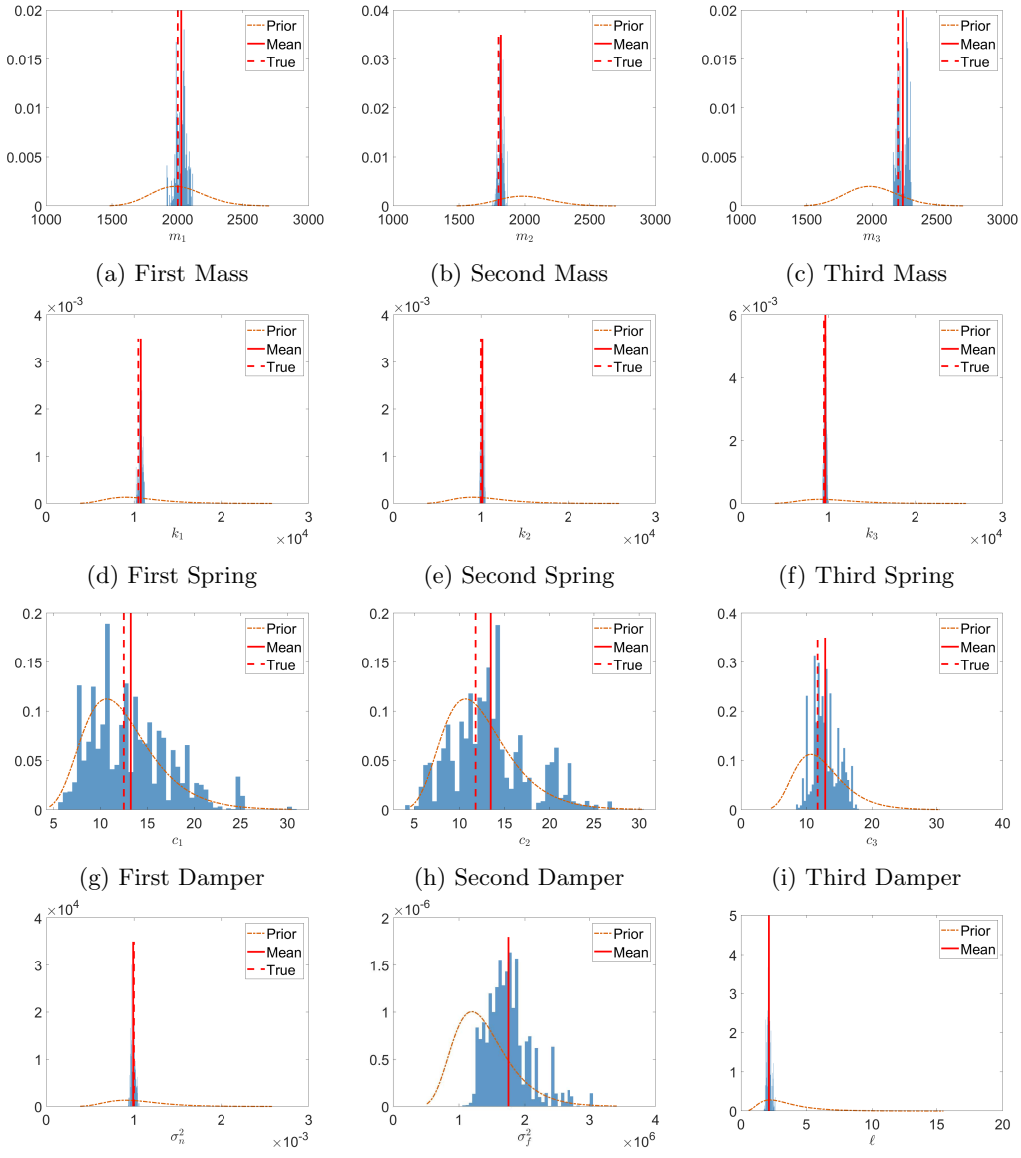
These results, along with Bayes rule, provide all the necessary theory to derive the Kalman filter and RTS smoother equations shown in Appendix A [24].

Appendix C. Partial Observation of Acceleration Measurements

The experiment conducted in Section 6 is repeated, except that the observation model is replaced with one where only the acceleration at the second and third mass are considered available.

With reference to the parameter estimation routine, it can be seen that the results are comparable to those obtained with displacement measurements, Fig. C.9. However, there is again significant difficulty in identifying the viscous damping parameters. It should also be noted that selection of the random walk variance to ensure convergence for this case required considerably more effort than for the displacement case. This difficultly should be expected as there is a reduction in available information for the system with the reduced set of observations. In general it is seen that this challenge results in an increased uncertainty regarding all parameters being identified. As was seen in the previous results the spring stiffness coefficients are the "easiest" to identify, there is the largest reduction in variance from the prior to the posterior and the posterior distributions have a mean very close to the ground truth. It should also be noted that in general, the quality of the MCMC results here are worse than those for the previous experiment. Visually this can be seen in the posterior histograms of the samples where the mixing in the chains has not resulted in the same tightly-packed unimodal posteriors seen previously.

Turning attention to the input estimation. There is a similar increase in the uncertainty over the smoothing distribution of the forcing, which is linked to both the increased parameter uncertainty and the reduced set of observed measurements. The recovered forcing signal on an independent test signal is shown in Fig. C.10. The NMSE for the means of the distributions are 12.83 on the training data (that used in the MCMC for parameter inference) and 13.06 on this independent test set. These point-wise error metrics on the mean of the forcing are significantly higher than those seen in the displacement case. However, investigation of the results shown in Figs. C.10 and C.11 reveals that the situation is slightly more nuanced. The forcing signal is seen to have a significantly increased level of uncertainty, this is likely due to the increase in parameter uncertainty and also the reduction in measurement information. If one were to draw samples from this distribution over the forcing, rather than considering only the mean, the measured signal is seen to lie well within the distribution. In the "zoomed " figure Fig. C.11, which focuses on the same section shown previously, the reconstructed force appears to have some offset from the mean which will lead to increased error. The reason for this is unclear and should be investigated further. However, if this estimation were to be carried forward into further analysis, e.g. fatigue analysis, then the reconstructed signal remains useful especially when considering if multiple realisations of the signal are used from the distribution to propagate uncertainty in the experienced forcing signal.



(a) First Mass (b) Second Mass (c) Third Mass

(d) First Spring (e) Second Spring (f) Third Spring

(g) First Damper (h) Second Damper (i) Third Damper

(j) Measurement Noise (k) GP Kernel Variance (l) GP Kernel Length Scale

Figure C.9: Posterior distributions over system parameters and Gaussian process hyperparameters for a system where the acceleration is observed at mass two and three only.

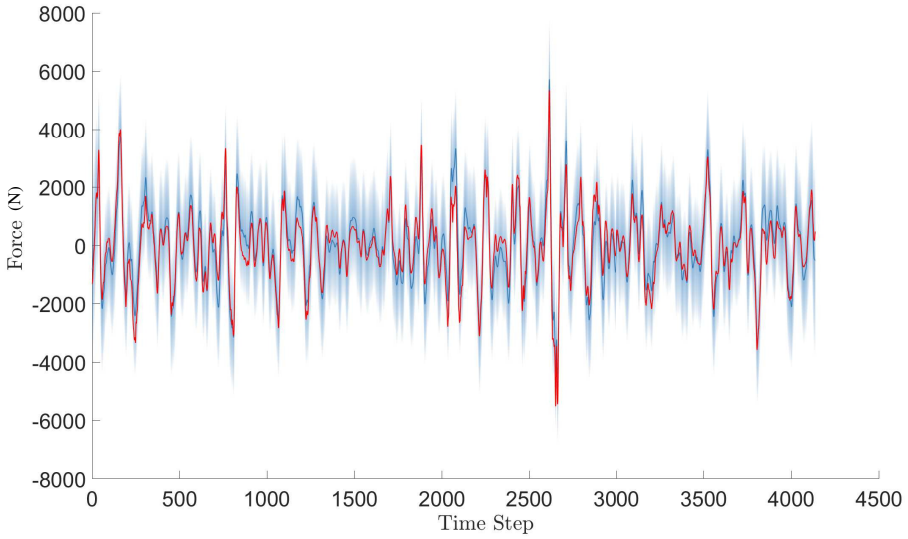


Figure C.10: Estimation of the input signal on an independent test set for the partially observed system

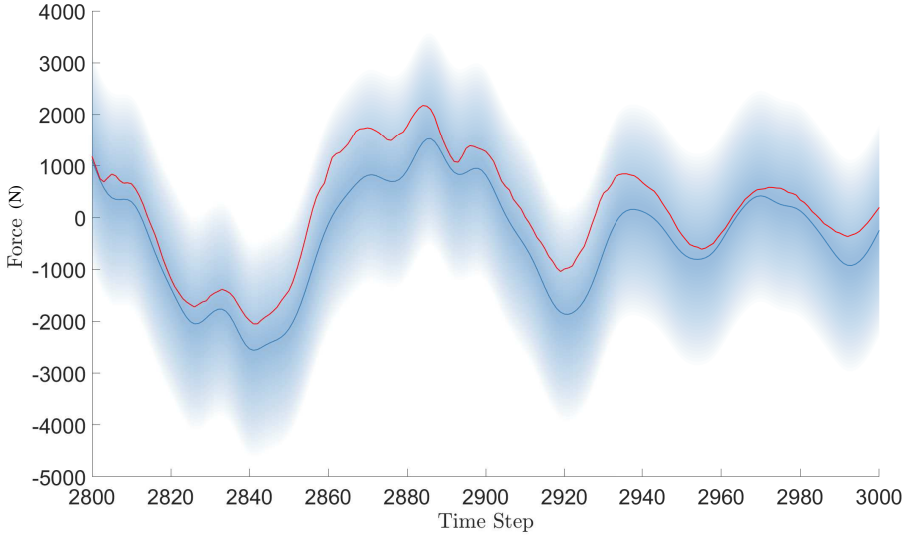


Figure C.11: Small section of the estimation of the input signal on an independent test set for the partially observed system.

The conclusion from this additional experiment is intuitive. While it is possible to still compute the quantities of interest, the parameter and forcing distributions, with a reduced measurement set, this will increase uncertainty in both the parameters and in the forcing. In addition to this, the inference procedure becomes practically more difficult as convergence of the MCMC scheme becomes harder to achieve. However, if this uncertainty is to be propagated forward this may not necessarily be an issue since the estimated distributions reflect the balance between the prior knowledge embedded in the system and the evidence seen in the reduced set of data being observed.