Journal of Machine Learning Research 1 (2000) 1-48 Submitted 4/00; Published 10/00

Efficient State-Space Inference of Periodic Latent Force Models

Steven Reece REECE@ROBOTS.OX.AC.UK
Stephen Roberts SJROB@ROBOTS.OX.AC.UK
Department of Engineering Science
University of Oxford
Parks Road
Oxford OX1 3PJ, UK

Siddhartha Ghosh SG2@ECS.SOTON.AC.UK
Alex Rogers ACR@ECS.SOTON.AC.UK
Electronics and Computer Science
University of Southampton
Southampton SO17 1BJ, UK

Nicholas R. Jennings NRJ@ECS.SOTON.AC.UK
Electronics and Computer Science
University of Southampton
Southampton SO17 1BJ, UK
and
Department of Computing and Information Technology
King Abdulaziz University
Saudi Arabia

Editor:

Abstract

Latent force models (LFM) are principled approaches to incorporating solutions to differential equations within non-parametric inference methods. Unfortunately, the development and application of LFMs can be inhibited by their computational cost, especially when closed-form solutions for the LFM are unavailable, as is the case in many real world problems where these latent forces exhibit periodic behaviour. Given this, we develop a new sparse representation of LFMs which considerably improves their computational efficiency, as well as broadening their applicability, in a principled way, to domains with periodic or near periodic latent forces. Our approach uses a linear basis model to approximate one generative model for each periodic force. We assume that the latent forces are generated from Gaussian process priors and develop a linear basis model which fully expresses these priors. We apply our approach to model the thermal dynamics of domestic buildings and show that it is effective at predicting day-ahead temperatures within the homes. We also apply our approach within queueing theory in which quasi-periodic arrival rates are modelled as latent forces. In both cases, we demonstrate that our approach can be implemented efficiently using state-space methods which encode the linear dynamic systems via LFMs. Further, we show that state estimates obtained using periodic latent force models can reduce the root mean squared error to 17% of that from non-periodic models and 27% of the nearest rival approach which is the resonator model (Särkkä et al., 2012; Hartikainen et al., 2012).

Keywords: latent force models, Gaussian processes, Kalman filter, kernel principle component analysis, queueing theory

1. INTRODUCTION

Latent force models (LFMs) have received considerable interest in the machine learning community as they combine underlying physical knowledge of a system with data driven models expressed as Bayesian non-parametric Gaussian process (GP) priors (see, for example, Alvarez et al., 2009; Hartikainen and Särkkä, 2010). In more detail, the physical process that generates the data is typically represented by one or more differential equations. These differential equations can then be accommodated within covariance functions along with the data driven priors. Doing so allows inferences to be drawn in regimes where data may be sparse or absent, where a purely data driven model will typically perform poorly. To date, such models have been applied in areas such as computational biology and understanding motion patterns (Alvarez et al., 2009, 2010).

Despite growing interest in LFMs, their real world applicability has been limited as inference using LFMs expressed directly through covariance functions can be computationally prohibitive on large data sets. It is well known that regression with GPs imposes high computational cost which scales as $\mathcal{O}(N^3T^3)$ during training, where N is the dimension of the data observed at each time point and T is the number of time points. However, it has also been shown that training LFMs using state-space methods can be considerably less computationally demanding (Rasmussen and Williams, 2006; Hartikainen and Särkkä, 2010) as state-space methods scale as $\mathcal{O}(N^3T)$. It is this computational saving that motivates the state-space approach to LFM inference in this paper.

The state-space approach to LFM inference advocated by Hartikainen and Särkkä (2010, 2011) augments the state vector so that Matérn and squared-exponential priors can be accommodated (although only approximately in the case of the squared-exponential). All the information encoded within the GP prior (that is, process smoothness, stationarity etc) is fully captured within their state-space representation. However, their approach assumes that the LFM kernel's inverse power spectrum can be represented by a power series in the frequency domain. Unfortunately, this requirement severely inhibits the applicability of their approach and, consequently, only a small repertoire of GP priors have been investigated within LFMs to date, namely, squared-exponential and Matérn kernels. Specifically, the state-space approach advocated by Hartikainen and Särkkä (2010) does not accommodate periodic kernels as we shall demonstrate in this paper. This is a key limitation as periodicity is common in many physical processes as we shall demonstrate in our empirical evaluation. Expressing our prior knowledge of the periodicity, as a GP prior, within the state-space approach is the key challenge problem addressed in this paper.

Thus, against this background, we describe a principled method for incorporating stationary periodic, non-stationary periodic and quasi-periodic Gaussian process priors within state-space approaches to LFM inference. Within our approach all LFM parameters can be inferred using Bayesian methods or maximum likelihood and thus we circumvent the need to set any of these parameters by hand. Further, to accommodate periodic and quasi-periodic models within LFMs we develop a novel state-space approach to inference. In particular, we propose to represent periodic and quasi-periodic driving forces, which are assumed

smooth, by linear basis models (LBMs) with eigenfunction basis functions derived using kernel principal component analysis (KPCA) in the temporal domain. These basis models, although parametric in form, are optimised so that their generative properties accurately approximate the driving force kernel. We will show that efficient inference can then be performed using a state-space approach by augmenting the state with additional variables which sparsely represent the periodic latent forces.

Our LBM approach to accommodating periodic kernels is inspired by the *resonator model* (Särkkä et al., 2012; Hartikainen et al., 2012) in which the periodic process is modelled as a superposition of *resonators*, each of which can be represented within the state-vector. Unfortunately, the resonator model, in its current form, does not encode all the underlying GP prior information of the periodic process as the resonator is not tailored to accommodate all the prior information encoded via the covariance function (see Section 4 for more detail). An alternative approach to modelling stationary kernels, including periodic kernels, is sparse spectrum Gaussian process regression (SSGPR) of Lázaro-Gredilla et al. (2010). This approach is similar in spirit to the resonator model in that it encodes stationary GP priors via basis functions (sinusoidal functions, in this case). However, unlike the resonator model, the SSGPR is able to encode the GP prior by reinterpreting the spectral density of a stationary GP kernel as a probability density function over frequency space. This pdf is then sampled using Monte Carlo to yield the frequencies of the sinusoidal basis functions. Unfortunately, this stochastic approach can often provide a poor approximation to the covariance function (see Section 5 for more detail).

We shall develop a LBM which captures all the information encoded within the GP prior and demonstrate its superior accuracy over the resonator model and the SSGPR. We shall also establish the close link between the resonator basis and the eigenfunction basis used in our approach and consequently, derive a novel method for tailoring the resonator basis to accommodate all the information encoded within the covariance function.

Our research is driven by two specific applications although the methods that we propose are of general applicability. Specifically, we apply our approach to the estimation and prediction of the behaviour of customer queues in call centres, based on flow models of queue dynamics represented as LFMs. The behaviour of queues is of general importance in several applications including communication networks (Wang et al., 1996), weather monitoring (Sims et al., 2005) and truck coordination at ports (Ji and Zhou, 2010). Accurate predictions of the customer queue arrival rates based on an underlying LFM is a key requirement for determining the number of call centre agents required at various times throughout the day. We also apply our approach to the estimation and prediction of the internal temperature within a home, based on thermal models of home heating systems represented as LFMs. Accurate predictions of the internal temperature based on an underlying LFM is a key component for predicting energy used in heating a home and, consequently, an integral part of many home energy saving systems (Bacher and Madsen, 2011). These applications demonstrate our approach under two different modelling conditions, the queue LFM is nonlinear whereas the thermal LFM is linear, while the queue application is a tracking application and regular measurements are available whereas the thermal application requires long term predictions (a day ahead) during which no measurements are available.

In more detail, telephone call centre managers are concerned with staffing and specifically, assigning the appropriate number of agents to guarantee that the customers' queueing

time does not prohibit sales (Feigin et al., 2006). Although there is significant literature on attempts to accurately model the dynamics of queues, it has failed to offer a method for inferring the highly quasi-periodic arrival rates from sparse measurements of the queue lengths (Wang et al., 1996). Determining such arrival rates is key to predicting queue lengths, and hence customer waiting times. These predictions help the call centre manager to plan staffing throughout the day to ensure an acceptable customer waiting time. We will demonstrate that our approach to modelling LFMs is capable of inferring these unknown arrival rates. Furthermore, although the dynamic system in this application is nonlinear and the arrival rate is quasi-periodic, it is still Markovian and, consequently, a state-space approach to inference is ideally suited to this application.

Energy saving in homes is a key issue as governments aim to reduce the carbon footprint of their countries. A significant amount of energy is expended in heating homes and home owners need to be encouraged to reduce their energy consumption and carbon emissions incurred through home heating (MacKay, 2009; DECC, 2009). Consequently, we apply our approach to the estimation and prediction of internal temperatures using thermal models of home heating systems. Our approach allows us to make day ahead predictions of the energy usage, which can then be fed back to the householder in real-time so that they can take appropriate mitigating actions to reduce their energy consumption. Home heating systems typically consist of a thermostat with a set-point that controls the activations of a gas or electrical boiler to ensure that the internal temperature follows the set-point. Although there is significant literature on attempts to accurately model the thermal dynamics of buildings, it has failed to take into account the daily human behaviours within their homes, which can have a significant impact on the energy signatures obtained from similar homes (Bacher and Madsen, 2011). For instance, during cold periods, a householder may deploy an additional heater or, in hot periods, open a window. Furthermore, the thermal dynamics of real homes are more complex in reality than existing thermal models suggest; sunlight through windows contributes to extra heat while open windows cause heat loss. Residual heat can also be retained by thermal blocks such as walls and ceilings that then re-radiate heat. Crucially, many of these heat sources are periodic in nature. For instance, an additional heater may be switched on every night during cold periods, whilst the diurnal sun cycle will contribute additional heat during the day. We will demonstrate that our approach is capable of inferring these unknown periodic heat sources. Again, the dynamic system in this application is linear and Markovian and, consequently, a state-space approach to inference is again ideally suited to this problem.

In undertaking this work, we advance the state of the art in the following ways:

- We offer the only principled approach to incorporating all Gaussian process prior models within a state-space approach to inference with LFMs. [1]

- We are the first to demonstrate that the eigenfunction model of Gaussian process priors out-performs an alternative approach to modelling periodic Gaussian process priors; namely, the sparse spectrum Gaussian process regression (SSGPR) approach developed by Lázaro-Gredilla et al. (2010).

1. What this paper does not aim to establish is the value of GP models per se over other models. The paper thus focusses on developing efficient, scalable representations and tools for performing GP inference.

- We demonstrate, for the first time, the close link between the eigenfunction model and the resonator model (Särkkä et al., 2012; Hartikainen et al., 2012; Solin and Särkkä, 2013). Consequently, we offer a novel mechanism for incorporating all information encoded within the latent force covariance function into the resonator model.

- We propose the only approach that is able to incorporate all types of periodic Gaussian process priors within a state-space approach to LFM inference. These priors include stationary periodic, non-stationary periodic and quasi-periodic covariance functions.

- We are the first to apply LFMs to queueing theory, specifically to the modelling of queue arrival rates. Through empirical evaluation, we show that for tracking the customer queue lengths in the call centre application, the RMSE of our approach using a quasi-periodic kernel model of the arrival rate can be 17% of that using the same approach with a non-periodic kernel model.

- We are the first to apply LFMs to the modelling of thermal dynamics within real homes, specifically to unknown physical thermal processes. We show that for day ahead predictions of temperature in homes, the RMSE of our approach is 45% of that obtained using the resonator model (Solin and Särkkä, 2013) when the latent forces exhibit quasi-periodic behaviour.

The structure of our paper is as follows: in Section 2 we review approaches to regression and time-series analysis using Gaussian processes and the Kalman filter. In Section 3 we review LFMs with a particular focus on periodic latent forces and then in Section 4 we present a critique of the existing state-space approaches to inference with LFMs. In Section 5 we present a novel approach to representing periodic LFMs by linear basis models. We critique the existing spectral models for representing periodic, stationary Gaussian process priors and argue that kernel principal component analysis is the most effective approach to inferring the Fourier basis for the corresponding LBMs. In Section 6 we extend our approach to representing quasi-periodic latent forces by linear basis models. Then, in Section 7 (with further details in Appendix A), we derive a state-space approach to inference with LFMs which accommodates both periodic and quasi-periodic forces via LBMs. In Section 8, we empirically demonstrate the utility of our approach in tracking the length of call centre customer queues in the presence of, initially, unknown arrival rates which are modelled as latent forces. In Section 9 we also apply our approach to predicting the internal temperature of homes in the presence of, a priori, unknown residual heat periodic latent forces. Furthermore, we demonstrate our approach on both single output and multi-output Gaussian process thermal models. We conclude in Section 10. Finally, in Appendix B we demonstrate the theoretical link between the eigenfunction basis used in our approach and the basis used within the resonator model. Consequently, we offer a novel method for encoding periodic latent force covariance functions within the resonator model.

2. A REVIEW OF GAUSSIAN PROCESS PRIORS AND INFERENCE

A Gaussian process (GP) is often thought of as a Gaussian distribution over functions (Rasmussen and Williams, 2006). A GP is fully described by its *mean function*, μ, and

covariance function, K. A draw, f, from a GP is traditionally written,

$$f \sim \mathcal{GP}(\mu, K) \ .$$

The value of the function, f, at inputs X is denoted $f(X)$. Similarly, the value of the mean function and covariance function at these inputs are denoted $\mu(X)$ and $K(X, X)$, respectively. The meaning of a GP becomes clear when we consider that, for any finite set of inputs, X, $f(X)$ is a draw from a multi-variate Gaussian, $f(X) \sim \mathcal{N}(\mu(X), K(X, X))$.

Suppose we have a set of training data,

$$D = \{(x_1, y_1), \ldots, (x_n, y_n)\} \ , \tag{1}$$

drawn from a function, f,

$$y_i = f(x_i) + \epsilon_i \ ,$$

where ϵ_i is a zero-mean Gaussian random variable with variance σ^2. For convenience both inputs and outputs are aggregated into sets $X = \{x_1, \ldots, x_n\}$ and $Y = \{y_1, \ldots, y_n\}$, respectively. The GP estimates the value of the function f at test inputs $X_* = \{x_{*1}, \ldots, x_{*m}\}$. The basic GP regression equations are given by,

$$\bar{f}_* = \mu(X_*) + K(X_*, X)[K(X, X) + \sigma^2 I]^{-1}(Y - \mu(X)) \ , \tag{2}$$

$$\mathrm{Var}(f_*) = K(X_*, X_*) - K(X_*, X)[K(X, X) + \sigma^2 I]^{-1} K(X_*, X)^T \ , \tag{3}$$

where I is the identity matrix, $\bar{f}_*$ is the posterior mean function at X_* and $\mathrm{Var}(f_*)$ is the posterior covariance (Rasmussen and Williams, 2006). The inversion operation present in Equations (2) and (3) is the source of the cubic computational complexity reported in the previous section.

The matrix $K(X, X)$ is the covariance of the Gaussian prior distribution over $f(X)$. The covariance matrix has elements,

$$K(x_i, x_j) = \mathrm{Cov}(f(x_i), f(x_j)) \ ,$$

where the term $K(X_*, X)$ is the covariance between the function, f, evaluated at the test inputs X_* and the training inputs X. The function K is alternatively called the *kernel* or the *covariance function*. There are many off-the-shelf kernels available (see, for example, Rasmussen and Williams (2006)) and appropriate kernels are chosen to model functions with requisite qualitative properties such as smoothness and stationarity. Further, basic kernels can be combined together to form more sophisticated kernels tailored to particular modelling needs. The mean function encodes our prior knowledge of the function mean. For ease of exposition we will assume that the mean function is zero a priori although the approaches to GP inference presented in later sections are not limited to this case.

The GP parameters θ (which includes σ and hyperparameters associated with the covariance function) can be inferred from the data through Bayes' rule,

$$p(\theta \mid Y, X) = \frac{p(Y \mid X, \theta)}{p(Y \mid X)} p(\theta) \ .$$

The parameters are usually given a vague prior distribution $p(\theta)$. In this paper, since our applications in Sections 8 and 9 exploit large data sets, we use maximum likelihood to infer the parameters and identify the assumed unique value for θ which maximises $p(Y \mid X, \theta)$. This approach is preferred over full Bayesian marginalisation (Bishop, 1999) as the preponderance of data in the applications we consider produces very tight posterior distributions over the parameters.

When the Gaussian process models a time series then the input variables, X, are values of time. We shall assume that increasing input indices correspond to sequential time stamps, $x_1 \leq x_2 \leq \ldots \leq x_{n-1} \leq x_n$. We are at liberty to deploy GP inference using Equations (2) and (3) to either *interpolate* the function $f(x_*)$ at x_* when $x_1 < x_* < x_n$ or *extrapolate* $f(x_*)$ when either $x_* < x_1$ or $x_* > x_n$. When measurements are obtained sequentially, extrapolation forward in time is termed *prediction* and the inference of $f(x_*)$ is termed *filtering*. Interpolation with sequential measurements is termed *smoothing*. Although both smoothing and filtering approaches have been developed for Gaussian process regression (Hartikainen and Särkkä, 2010), we shall be concerned with filtering only. However, the eigenfunction models for periodic Gaussian processes developed in this paper can also be used for smoothing.

In the next section we review the latent force model (LFM) which is a principled approach to incorporating solutions to differential equations within Gaussian process inference methods.

3. LATENT FORCE MODELS

In this section we present a brief introduction to latent force models and describe their practical limitations. Specifically, we consider dynamic processes which can be described by a set of E coupled, stochastic, linear, first order differential equations,

$$\frac{dz_q(t)}{dt} = \sum_{e=1}^{E} F_{e,q} z_e(t) + \sum_{r=1}^{R} L_{r,q} u_r(t) \ ,$$

where q and e index each variable z, R is the number of latent forces and r indexes each latent force u, and L and F are coefficients of the system. For example, in our home heating application (described in detail in Section 9), $z_1(t)$ models the internal temperature of a home, $z_2(t)$ the ambient temperature immediately outside the home, $u_1(t)$ is the heater output from a known proportional controller and $u_2(t)$ is an unknown residual force. In this application, we assume $u_2(t)$ is periodic as it is used to model solar warming, some habitual human behaviour and the thermal lags in the heating system. The resulting differential equations can be written as,

$$\frac{d\mathbf{z}(t)}{dt} = \mathbf{F}\ \mathbf{z}(t) + \mathbf{L}\mathbf{u}(t) \ , \tag{4}$$

where $\mathbf{u}(t)$ is a vector of R independent *driving forces* (also called the *latent forces*). We distinguish non-periodic latent forces, `np`, and periodic latent forces, `p`, as they will be modelled differently in our approach. Non-periodic forces will be modelled using the existing approach advocated in Hartikainen and Särkkä (2010), which is reviewed in Section 4, and

periodic forces will be modelled using our novel linear basis approach presented in Section 5. In Equation (4) the $E \times E$ matrix $\mathbf{F}$ and the $E \times R$ matrix $\mathbf{L}$ are non-random coefficients that link the latent forces to the dynamic processes. Although we deal with first order differential equations only, all higher order differential equations can be converted to a set of coupled first order equations (Hartikainen and Särkkä, 2010).

Following Alvarez et al. (2009) and Hartikainen and Särkkä (2010, 2011) we assume that the latent forces, $\mathbf{u}$, are independent draws from Gaussian processes, $u_i \sim \mathcal{GP}(0, K_i)$ where K_i is the GP covariance function (Rasmussen and Williams, 2006) for force u_i. Consequently, the covariance for $\mathbf{z}$ at any times t and t' can be evaluated as follows,

$$E[(\mathbf{z}(t) - \bar{\mathbf{z}}(t))(\mathbf{z}(t') - \bar{\mathbf{z}}(t'))^T] = \mathbf{\Phi}(t_0, t)\mathbf{P}_z^0\mathbf{\Phi}(t_0, t')^T + \Gamma(t_0, t, t') \ , \tag{5}$$

where $\mathbf{\Phi}(t_0, t)$ denotes the matrix exponential, $\mathbf{\Phi}(t_0, t) = \exp(\mathbf{F}(t - t_0))$ expressed in Alvarez et al. (2009), $\bar{\mathbf{z}}(t) = E[\mathbf{z}(t)]$, [2]

$$\Gamma(t_0, t, t') = \int_{t_0}^{t}\int_{t_0}^{t'} \mathbf{\Phi}(s, t)\mathbf{L}\mathbf{K}(s, s')\mathbf{L}^T\mathbf{\Phi}(s', t')^T \ ds \ ds' \ ,$$

$\mathbf{P}_z^0$ is the state covariance at time t_0, $\mathbf{P}_z^0 = E[(\mathbf{z}(t_0) - \bar{\mathbf{z}}(t_0))(\mathbf{z}(t_0) - \bar{\mathbf{z}}(t_0))^T]$ and $\mathbf{K}(s, s')$ is the diagonal matrix $\mathbf{K}(s, s') = \text{diag}(K_1(s, s'), \ldots, K_R(s, s'))$. Since $\mathbf{z}(t)$ is a vector and defined for any times t and t' then $E[(\mathbf{z}(t) - \bar{\mathbf{z}}(t))(\mathbf{z}(t') - \bar{\mathbf{z}}(t'))^T]$ is a multi-output Gaussian process covariance function. A kernel for covariances between the target, $\mathbf{z}$, and the latent forces, $\mathbf{u}$, can also be derived. Inference with these kernels is then undertaken directly using Equations (2) and (3).

Unfortunately, a naïve implementation of LFM inference using Equations (2) and (3) and covariance functions derived using Equation (5) can be computationally prohibitive. As we have already pointed out, this approach can be computationally expensive due to the need to invert prohibitively large covariance matrices. To mitigate computational intensive matrix inversion in the GP equations, various sparse solutions have been proposed (see, for example, Williams and Seeger, 2001; Snelson and Ghahramani, 2006; Lázaro-Gredilla et al., 2010) and an early review of some of these methods is presented in Quiñonero Candela and Rasmussen (2005). Unfortunately, the spectral decomposition approach of Lázaro-Gredilla et al. (2010) is sub-optimal in that it randomly assigns the components of a sparse spectral representation and this limitation is explored in detail in Section 5. The Nyström method for approximating eigenfunctions is used in Williams and Seeger (2001) to derive a sparse approximation for the kernel which can then be used to improve the computational efficiency of the GP inference Equations (2) and (3). Unfortunately, this approximate kernel is not used consistently throughout the GP equations and this can lead incorrectly to negative predicted variances.

The pseudo-input (also called *inducing inputs*) approach (Snelson and Ghahramani, 2006; Quiñonero Candela and Rasmussen, 2005) is a successful method for reducing the number of input samples used within GP inference without significantly losing information encoded within the full data set. In essence, densely packed samples are summarised around sparsely distributed inducing points. Pseudo-inputs have been successfully deployed within sparse approximations of dependent output Gaussian processes (Alvarez and Lawrence,

2. All integrals in this paper should be interpreted as Itô integrals.

2008, 2011). Pseudo-inputs have recently been introduced to GP time-series inference and applied to problems which exploit differential equations of the physical process via the latent force model (Alvarez et al., 2011). In Alvarez et al. (2011) the latent force is expressed at pseudo-inputs and then convolved with a smooth function to interpolate between the pseudo-inputs. However, although inducing inputs can reduce the sampling rate and summarise local information, they still have to be liberally distributed over the entire time sequence. We may assume, for simplicity, the pseudo-inputs are evenly spread over time and, therefore, the number of pseudo-inputs, P, would have to increase linearly with the number of observations (although with a rate considerably lower than the observation sampling rate). Unfortunately, the computational complexity of GP inference with pseudo-inputs is $\mathcal{O}(TP^2)$ where T is the number of observations (Alvarez and Lawrence, 2008). Thus, although pseudo-inputs are able to improve the efficiency of GP inference to some extent, for time series analysis their computational cost is still cubic in the number of measurements and this can be computationally prohibitive.

In the next section we describe a state-space reformulation of the LFM. The state-space approach has the advantage that it has a computational complexity for inferring the target process, $\mathbf{z}$, which is $\mathcal{O}(T)$ but at the expense of representing the target process with extra *state* variables.

4. STATE-SPACE APPROACHES TO LATENT FORCE INFERENCE

In this section we review the current state-space approach to inference with LFMs (Hartikainen and Särkkä, 2010) and show how some covariance functions can be represented exactly in state-space. Unfortunately, we shall also demonstrate that periodic kernels cannot be incorporated into LFMs using the approach advocated by Hartikainen and Särkkä (2010). To address this key issue, we will then propose to approximate a periodic covariance function with a sparse linear basis model. This will allow us to represent periodic behaviour within a LFM efficiently and also incorporate information encoded within the periodic kernel prior. Our work is inspired by, and can be seen as, an extension of the resonator model (Särkkä et al., 2012; Hartikainen et al., 2012), which is an alternative linear basis model that allows periodic processes to be modelled within the state-space approach. Our LBM approach, described in Section 5, builds on the resonator model and extends it by incorporating the prior information encoded within the latent force covariance function.

When the target processes, $\mathbf{z}$ as per Equation (4), can be expressed in Markov form, we can avoid the need to invert large covariance matrices and also avoid the need to evaluate Equation (5) over long time intervals, $[t_0,\ t]$, by using the more efficient state-space inference approach advocated by Hartikainen and Särkkä (2010) and in this paper. The temporal computational complexity of the state-space approach is $\mathcal{O}(T)$ as we integrate over short time intervals, $[t_0,\ t]$, and then reconstruct long term integrations by conflating the local integrations via the Kalman filter. This is an alternative approach to that advocated by Alvarez et al. (2009) in which we integrate the differential equations, as per Equation (5), over long intervals, $[t_0,\ t]$, and then regress using Equations (2) and (3). Both approaches are mathematically equivalent in that they produce identical inferences when they are applied to the same differential model, latent force covariance functions and data.

The Kalman filter is a state-space tool for time series estimation with Gaussian processes (Kalman et al., 1960). The *Kalman smoother* is also available for interpolation with sequential data. The Kalman filter is a state-space inference tool which summarises all information about the process, f, at time x via a *state* description. The advantage of the Kalman filter is that any process f_* at any future time x_* can be inferred from the current state without any need to refer to the process history. The state at any time x is captured by a finite set of Gaussian distributed *state variables*, $\mathbf{U}$, and we assume that f is a linear function of the state variables. In Hartikainen and Särkkä (2010) the state variables corresponding to each latent force f are the function f and its derivatives. In our approach the state variables corresponding to each periodic latent force will be the eigenfunctions of the periodic covariance function. The key advantage of the Kalman filter is that its computational complexity is linear in the amount of data from a single output time-series. Contrast this with the standard Gaussian process approach, as per Equations (2) and (3), which require the inversion of a covariance matrix and thus, have a computational complexity which is cubic in the amount of data. [3]

To illustrate the state-space approach consider a single non-periodic latent force, $u_r(t)$, indexed by r, in Equation (4). We assume that this force is drawn from a Gaussian process,

$$u_r \sim \mathcal{GP}(0, K_r) \ ,$$

where K_r is a stationary kernel. In Hartikainen and Särkkä (2010) the authors demonstrate that a large range of stationary Gaussian process kernels, K_r, representing the latent force prior can be transformed into multivariate linear time-invariant (LTI) stochastic differential equations of the form,

$$\frac{d\mathbf{U}_r(t)}{dt} = \mathbf{F}_r \ \mathbf{U}_r(t) + \mathbf{W}_r \ \omega_r(t) \ , \tag{6}$$

where $\mathbf{U}_r(t) = (u_r(t), \ \frac{du_r(t)}{dt}, \ \cdots, \frac{d^{p_r-1}u_r(t)}{dt^{p_r-1}})^T$ and,

$$\mathbf{F}_r = \begin{pmatrix} 0 & 1 & & \\ & \ddots & \ddots & \\ & & 0 & 1 \\ -c_r^0 & \cdots & -c_r^{p_r-2} & -c_r^{p_r-1} \end{pmatrix} \ , \qquad \mathbf{W}_r = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix} \ , \tag{7}$$

where c are coefficients which can be set using spectral analysis of the kernel as per Hartikainen and Särkkä (2010). The force, $u_r(t)$, can be recovered from $\mathbf{U}_r(t)$ using the indicator vector $\mathbf{\Delta}_r = (1, 0, \ldots, 0)$ where,

$$u_r(t) = \mathbf{\Delta}_r \mathbf{U}_r(t) \ .$$

By choosing the coefficients $c_r^0, \ldots, c_r^{p_r-1}$ in Equation (7), the spectral density of the white noise process $\omega_r(t)$ in Equation (6) and the dimensionality p_r of $\mathbf{U}_r(t)$ appropriately, the covariance of $u_r(t)$, corresponding to the dynamic model, can be chosen to correspond to the GP prior K_r. The differential equations expressed in Equation (6) can then be integrated

3. The Kalman filter has a cubic computational complexity in the number of measured processes for multi-output Gaussian processes. We shall clarify the computational complexity of Kalman filter models for multi-output GPs in Section 7 and investigate an application of multi-output GPs in Section 9.

into the LFM to form the augmented dynamic model expressed later in Equation (12). The coefficients $c_r^0, \ldots, c_r^{p_r-1}$ are found by initially taking the Fourier transform of both sides of Equation (6). The coefficients can then be expressed in terms of the spectral density of the latent force kernel, K_r, provided that its spectral density, $S_r(\varpi)$, can be written as a rational function of ϖ^2,

$$S_r(\varpi) = \frac{(\text{constant})}{(\text{polynomial in } \varpi^2)} \ . \tag{8}$$

The *inverse* power spectrum is then approximated by a polynomial series from which the transfer function of an equivalent stable Markov process for the kernel can be inferred along with the corresponding spectral density of the white noise process. The stochastic differential equation coefficients are then calculated from the transfer function. For example, for the first-order Matérn kernel,

$$K_r(t,t') = \sigma_r^2 \exp\left(-\frac{|t-t'|}{l_r}\right) \ , \tag{9}$$

with output scale σ_r and input scale l_r, $u_r \sim \mathcal{GP}(0, K_r)$ can be represented by Equation (6) with $\mathbf{U}_r(t) = u_r$, $\mathbf{W}_r = 1$ and,

$$\mathbf{F}_r = -1/l_r \ . \tag{10}$$

The spectral density, λ_r, of the white noise process, ω_r, is,

$$\lambda_r = \frac{2\sigma_r^2\sqrt{\pi}}{l_r\ \Gamma(0.5)} \ , \tag{11}$$

and Γ is the Gamma function (Hartikainen and Särkkä, 2010).

Now, by augmenting the state vector, $\mathbf{z}$ in Equation (4) appropriately with the non-periodic forces $\mathbf{U}_r(t)$ and their derivatives Hartikainen and Särkkä (2011) demonstrate that the dynamic equation can be rewritten as a joint stochastic differential model,

$$\frac{d\mathbf{z}_a(t)}{dt} = \mathbf{F}_a\ \mathbf{z}_a(t) + \mathbf{L}_a\omega_a(t) \ , \tag{12}$$

where

$$\begin{aligned} \mathbf{z}_a(t) &= (\mathbf{z}(t),\ \mathbf{U}_1(t),\ \ldots\ ,\ \mathbf{U}_R(t))^T \ , \\ \mathbf{F}_a &= \begin{pmatrix} \mathbf{F} & \mathbf{L}\mathbf{S}_1\mathbf{\Delta}_1 & \ldots & \mathbf{L}\mathbf{S}_R\mathbf{\Delta}_R \\ \mathbf{0} & \mathbf{F}_1 & \ldots & \mathbf{0} \\ & & \vdots & \\ \mathbf{0} & \mathbf{0} & \ldots & \mathbf{F}_R \end{pmatrix} \ , \end{aligned}$$

R is the number of latent forces, $\mathbf{S}_r = (0, \ldots, 1, \ldots, 0)$ is the indicator vector which extracts the rth column of $\mathbf{L}$ corresponding to the rth force, u_r, and $\omega_a(t)$ is the appropriate scalar process noise,

$$\omega_a(t) = (0,\ \omega_1(t),\ \ldots\ ,\ \omega_R)^T \ , \tag{13}$$

$$\mathbf{L}_a = \text{blockdiag}(\mathbf{0},\ \mathbf{W}_1,\ \ldots\ ,\mathbf{W}_R) \ . \tag{14}$$

These differential equations have the solution,

$$\mathbf{z}_a(t) = \mathbf{\Phi}(t_0, t)\mathbf{z}_a(t_0) + \mathbf{q}_a(t_0, t) \ ,$$

where, again, $\mathbf{\Phi}(t_0, t)$ denotes the matrix exponential, $\mathbf{\Phi}(t_0, t) = \exp(\mathbf{F}_a(t - t_0))$ expressed in Alvarez et al. (2009). The process noise vector, $\mathbf{q}_a(t_0, t)$, is required to accommodate the Matérn or SE latent forces within the discrete time dynamic model, $\mathbf{q}_a(t_0, t) \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_a(t_0, t))$ where,

$$\mathbf{Q}_a(t_0, t) = \int_{t_0}^{t} \mathbf{\Phi}(s, t)\mathbf{L}_a\Lambda_a\mathbf{L}_a^T\mathbf{\Phi}(s, t)^T ds \ ,$$

and Λ_a is a diagonal matrix,

$$\Lambda_a = \text{diag}(0, \lambda_1, \ldots, \lambda_R) \ , \tag{15}$$

where λ_r is the spectral density of the white noise process corresponding to the Matérn or SE process, K_r (Hartikainen and Särkkä, 2010).

We now briefly describe the reasons why this spectral analysis approach advocated by Hartikainen and Särkkä (2010, 2011) cannot be immediately applied to periodic kernels. For illustrative purposes we shall investigate the commonly used squared-exponential periodic kernel,

$$K_{\text{SE}}(t, t') = \exp\left(-\frac{\sin\left(\frac{\pi(t-t')}{D}\right)^2}{l^2}\right) \ , \tag{16}$$

with input scale $l = 3$, an implicit output scale of unity and period $D = 0.7$, although our analysis and conclusions apply to all periodic kernels, in general. Unfortunately, as is shown in the left panel of Figure 1, the power spectrum for this periodic kernel is a weighted sum of Dirac delta functions, each delta function identifying a sinusoidal mode. The inverse of the power spectrum is highly nonlinear and not amenable to the polynomial series approximations expressed in Equation (8). The left panel also shows the best (in a least squares sense) polynomial fit to the inverse spectrum. The polynomial coefficients are shown in the central panel and very little weight is assigned to higher order frequencies. Now, using the approach advocated in Hartikainen and Särkkä (2010) we can infer the covariance function corresponding to this polynomial approximation of the inverse spectrum. This covariance function and the true periodic covariance function are shown in the right panel of Figure 1. It is clear that the covariance function obtained using Hartikainen and Särkkä (2010) is a poor representation of the true periodic kernel.

So, it is not possible to formulate all periodic latent forces via Equation (6). However, by approximating the latent force as a linear sum of basis functions, such that each basis function, ϕ, can be formulated via Equation (6),

$$u_r(t) = \sum_j a_{rj}\phi_j(t) \ , \tag{17}$$

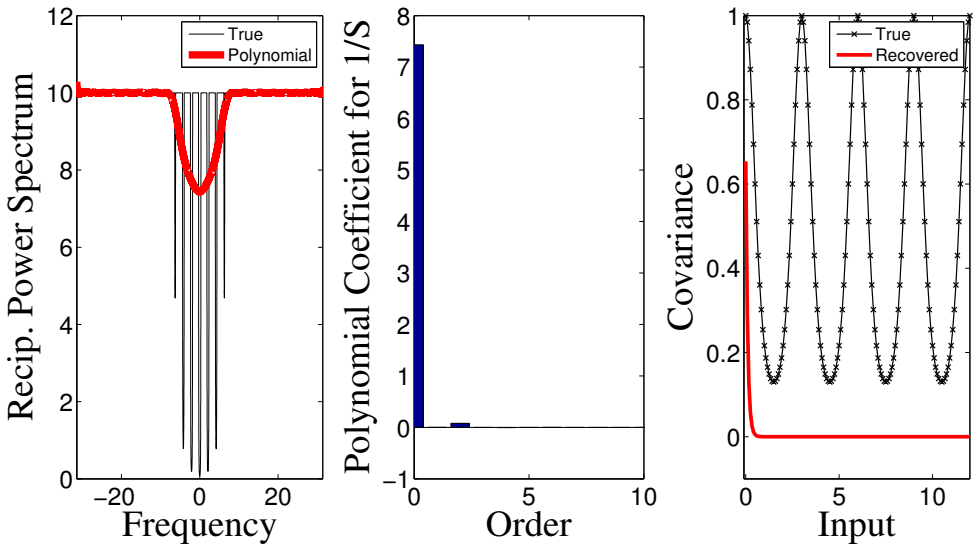


Figure 1: Spectral analysis of a periodic covariance function. The left panel shows inverse power spectrum for a periodic squared-exponential kernel (thin line) and its polynomial approximation (thick line). The central panel shows the coefficients of the polynomial approximation. The right panel shows the true covariance function (crossed line) and its approximation (solid line) recovered from the polynomial representation of the inverse power spectrum.

then it is possible to represent the periodic latent force within the KF. In essence, the latent force, u_r, is decomposed into a weighted sum of basis latent forces, $\{\phi_j\}$, such that each ϕ satisfies Equation (6). This is the approach of Särkkä et al. (2012) for representing both stationary and quasi-periodic latent forces via their *resonator model*. In Hartikainen et al. (2012), the *resonator*, ϕ_r, is chosen to be a Fourier basis, $\phi_r(t) = \cos(f_r t)$ or $\phi_r(t) = \sin(f_r t)$. The resonator can be represented by Equation (6) as a state comprising the instantaneous resonator value, $\phi_r(t)$, and its derivative, $\dot{\phi}_r(t)$ thus, $\mathbf{U}_r(t) = (\phi_r(t),\ \dot{\phi}_r(t))^T$. The corresponding SDE has $\mathbf{F}_r = [0 \quad 1\ ;\ -f_r^2 \quad 0]$ and $\mathbf{W}_r = 0$. The Fourier basis is particularly useful for modelling stationary covariance functions.

In Hartikainen et al. (2012) quasi-periodic latent forces were implemented as a superposition,

$$u(t) = \sum_j \psi_j(t)\ , \tag{18}$$

of resonators, ψ, of the form,

$$\frac{d^2\psi_j(t)}{dt^2} = -(2\pi f_j(t))^2\psi_j(t) + \omega_j(t)\ , \tag{19}$$

where ω is a white noise component. Crucially, the resonator *frequencies*, f, are time variant and this supports non-stationary and quasi-periodic forces. This model is very flexible and

both periodic and quasi-periodic processes can be expressed using the resonator model (as detailed in Appendix B). However, currently no mechanism has been proposed to incorporate prior information encoded in periodic GP kernels within the resonator model. Further, inferring the parameters and the frequency profiles, $f(t)$, for each resonator can be prohibitively computationally expensive (as we demonstrate in Appendix B). Despite these shortcomings there is a very close connection between the resonator model for periodic latent forces and the eigenfunction approach proposed in this paper. This connection is explored in detail in Appendix B in which we assert that the eigenfunction basis is an instance of the resonator basis for perfectly periodic covariance functions. We subsequently demonstrate how the eigenfunction approach can both inform the resonator model of the GP prior and also simplify the inference of the resonator model parameters including the frequency profile. Further, we show that the optimal minimum mean-square resonator model is an alternative way of representing the corresponding eigenfunction basis within the Kalman filter.

In the original implementation of the resonator model (Särkkä et al., 2012) the model parameters were set by hand. Recently, a new variation of the resonator model has been proposed in which the most likely model parameters are learned from the data (Solin and Särkkä, 2013). In this version the resonator is the solution to the time invariant second order differential equation,

$$\frac{d^2\psi_j(t)}{dt^2} = A_j\psi_j(t) + B_j\frac{d\psi_j(t)}{dt} + \omega_j(t) \ , \tag{20}$$

where A and B are constant coefficients. We note that this variation of the resonator model is a special case of the original resonator model with a frequency profile $f_j(t) = \frac{i}{2\pi}\sqrt{A_j + B_j\frac{1}{\psi_j(t)}\frac{d\psi_j(t)}{dt}}$ in Equation (19). To model quasi-periodic processes Equation (20) comprises a *decay* term via the first derivative of the resonator function. This new model is computationally efficient as it imposes constant coefficients unlike the original resonator model in Särkkä et al. (2012). However, the computational efficiency of the model in Equation (20), gained by losing the requirement to infer a frequency profile for each resonator, is at the expense of the model's flexibility. We compare the resonator model in Equation (20) with our eigenfunction approach on a real world application in Section 9.

In preparation for the approach advocated in this paper, in which we also represent the periodic kernel via a linear basis model, the following section compares the two key alternative approaches to directly inferring linear basis models from Gaussian Process kernels, namely the sparse spectrum Gaussian process regression (SSGPR, Lázaro-Gredilla et al., 2010) and kernel principal component analysis (Schölkopf and Müller, 1998).

5. REPRESENTING PERIODIC LATENT FORCES WITH LINEAR BASIS MODELS

In this section, we exploit linear basis models and propose a novel approach to representing periodic latent force GP kernels. Our aim is to derive a sparse representation for periodic kernels so that they can be accommodated within a state-space formulation of the LFM. Linear basis models (LBMs) have a long history in machine learning. In particular, special cases of them include kernel density estimators (Parzen, 1962) and the Relevance Vector Machine (Tipping, 2001). There are two key advantages to representing periodic kernels

using a sparse basis model: firstly, they can approximate the kernel using a weighted sum over a finite set of functions. As we will see, for relatively smooth kernels the number of basis functions can be small. The second advantage, as we will show in Section 7, is that the LBM representation is amenable to inference using computationally efficient state-space methods. We exploit the Nyström approximation as opposed to other sparse approximations (such as the sparse spectrum Gaussian process regression (SSGPR) method of Lázaro-Gredilla et al. (2010)) as, we will see, the eigenfunctions of the kernel form the most efficient basis for the corresponding driving forces. This approximation will accommodate both the prior information about the driving forces (encoded in the kernel) within a state-space approach and also provide a means to learn these driving forces from data using iterative state-space methods. Approximating Gaussian process priors via the Nyström method is not new (see, for example, Williams and Seeger, 2001). However, using this to accommodate periodic and quasi-periodic latent forces within LFMs is novel.

In order to develop our LBM for latent forces we shall first investigate current approaches to sparse representations of stationary covariance functions and then demonstrate that one of these approaches, namely the eigenfunction approach, generalises to non-stationary covariance functions. Bochner's theorem asserts that all stationary covariance functions can be expressed as the Fourier transform of their corresponding spectral densities (where the spectral density exists. See, for example, Rasmussen and Williams, 2006). Furthermore, in the stationary case, the Fourier basis is the eigenfunctions of the covariance function. There has been a long history of research into the spectral analysis of stationary Gaussian process kernels (see, for example, Bengio et al., 2004). However, only recently has the Fourier basis been investigated in the context of latent force models. To date, two approaches have been proposed to incorporate knowledge of all stationary kernels, including periodic kernels, within the linear basis representation via spectral analysis: the SSGPR (Lázaro-Gredilla et al., 2010) and the KPCA (Drineas and Mahoney, 2005) method. The key advantage of these approaches is that the basis frequencies can be calculated from the prior latent force kernel. These approaches are described and compared next.

The SSGPR (Lázaro-Gredilla et al., 2010) approach reinterprets the spectral density of a stationary GP kernel as the probability density function over frequency space. This pdf is then sampled using Monte Carlo to yield the frequencies of the sinusoidal basis functions of the LBM. [4] The advantage of this approach is that a sparse set of sinusoidal basis functions is identified such that the most significant frequencies of these sinusoidal basis functions have the greatest probability of being chosen. The phase of each basis function is then inferred from the data. The disadvantage of this approach is it can often provide a poor approximation to the covariance function as we will demonstrate shortly in Figure 3.

An alternative approach to the SSGPR is KPCA which effectively intelligently samples the most informative frequencies within the spectral density. Mercer's theorem (Mercer, 1909) allows us to represent each periodic latent force, $u(t)$, at arbitrary inputs, t, via an infinite set of basis functions, ϕ_j,

$$u(t) = \sum_{j=1}^{\infty} a_j \phi_j(t) \ , \tag{21}$$

4. In their code, available at `http://www.tsc.uc3m.es/~miguel/downloads.php`, the authors try several frequency initialisations and use the best one.

where $\{a_j\}$ are the model *weights* which are independently drawn from a Gaussian,

$$a_j \sim \mathcal{N}(0, \mu_j^\phi) \ , \tag{22}$$

where μ_j^ϕ is the variance of a_j. For any choice of probability density function, p, there exists an orthonormal basis, $\{\phi\}$, such that,

$$\int \phi_i(t)\phi_j(t)p(t)dt = \begin{cases} 1 & \text{if } \ i = j \ , \\ 0 & \text{otherwise} \ . \end{cases}$$

Furthermore, the latent force prior, $K(t,t') = E[u(t)u(t')]$, can be expressed as,

$$K(t,t') = \sum_{j=1}^{\infty} \mu_j^\phi \phi_j(t)\phi_j(t') \ , \tag{23}$$

where, ϕ_j are the *eigenfunctions* of the kernel, K, under p such that,

$$\int K(t,t')\phi_j(t')p(t')dt' = \mu_j^\phi \phi_j(t) \ , \tag{24}$$

and the variance, μ_j^ϕ, is also an *eigenvalue* of the kernel.

Of course, it is not feasible to actually use an infinite basis. Thus, we approximate the infinite sum in Equation (21) by a finite sum over a subset of significant eigenfunctions which have the J most significant eigenvalues, μ^ϕ,

$$u(t) \approx \sum_{j=1}^{J} a_j \phi_j(t) \ . \tag{25}$$

Fortunately, kernel principal component analysis (KPCA) allows us to identify the most significant J eigenfunctions a priori as well as compute their form approximately (Schölkopf and Müller, 1998).

The role of p, in Equation (24), is to weight the values of time t. We are free to choose the probability density function, $p(t)$, as we wish. For stationary covariance functions, a uniform pdf is appropriate as it weights each time instance, t, equally. To evaluate the integral in Equation (24) we use a quadrature-based method and N equally spaced quadrature points, S, of t, where $S = \{s_1, \ldots, s_N\}$ (see, for example, Shawe-Taylor et al., 2005),

$$\int K(t,t')\phi_j(t')p(t')dt' \approx \frac{1}{N}\sum_{i=1}^{N} K(t,s_i)\phi_j(s_i) \ . \tag{26}$$

The points, S, are also used to construct an $N \times N$ covariance matrix, G, called the *Gram matrix*, where,

$$G_{ij} = K(s_i, s_j) \ . \tag{27}$$

The Nyström approach is then used to derive approximate eigenfunctions of K using the eigenvectors, $\mathbf{v}$, and eigenvalues, μ, of the Gram matrix (Drineas and Mahoney, 2005). We

denote the Nyström approximation for ϕ_j with uniform pdf p as $\tilde{\phi}_j$. For each eigenvector, $\mathbf{v}_j$,

$$\tilde{\phi}_j(t) = \frac{\sqrt{N}}{\mu_j} K(t, S)\mathbf{v}_j \ . \tag{28}$$

Since $\{\mathbf{v}_j\}$ are orthonormal then $\{\tilde{\phi}_j\}$ are orthogonal. Now, substituting the approximation for ϕ into Equation (25),

$$u(t) \approx \sum_{j=1}^{J} a_j \tilde{\phi}_j(t) \ . \tag{29}$$

By forming the covariance between $u(t)$ and $u(t')$ we can derive the following relationship between the latent force prior, the approximate eigenfunctions and the variances μ_j^{ϕ} of the model weights, a_j,

$$K(t, t') \approx \sum_{j=1}^{J} \mu_j^{\phi} \tilde{\phi}_j(t) \tilde{\phi}_j(t') \ , \tag{30}$$

where $\mu_j^{\phi} \approx \mu_j / N$, is the scaled Gram matrix eigenvalue (Williams and Seeger, 2001).

As we can compare the covariance function, K, with the corresponding Nyström covariance function approximation, as per Equation (30), then the sample set, S, can be chosen a priori to provide a comprehensive representation of the kernel K. Furthermore, as $N \to \infty$ then $\tilde{\phi}_j \to \phi_j$. Finally, although the eigenfunction LBM is a parametric model, the eigenfunctions accurately reproduce the periodic GP prior across an entire period and undesirable extrapolation errors often associated with spatially degenerate LBMs are alleviated here (Rasmussen and Williams, 2006).

Throughout this paper the LBMs will comprise the most significant eigenfunctions according to the following definition,

Definition 1 *An eigenfunction is* significant *if its eigenvalue is more than a pre-defined fraction γ of the maximum eigenvalue.*

We have found that $\gamma = 1/100$ is a robust choice for the applications in Sections 8 and 9 in which fewer than 30 basis functions are required to model the latent forces.

To demonstrate the eigenfunction approach to representing Gaussian process priors via a finite basis, Figure 2(a) shows example eigenfunctions for a stationary periodic Matérn process. The Matérn kernel is defined (Rasmussen and Williams, 2006),

$$\text{Matérn}(\tau, \nu, \sigma, l) = \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{\sqrt{2\nu}}{l} \tau \right)^{\nu} \check{K}_{\nu} \left(\frac{\sqrt{2\nu}}{l} \tau \right) \ , \tag{31}$$

where $\tau \geq 0$, Γ and $\check{K}_{\nu}$ are the gamma and modified Bessel functions, respectively, ν indicates the order, σ is the output scale which governs the amplitude of the kernel and

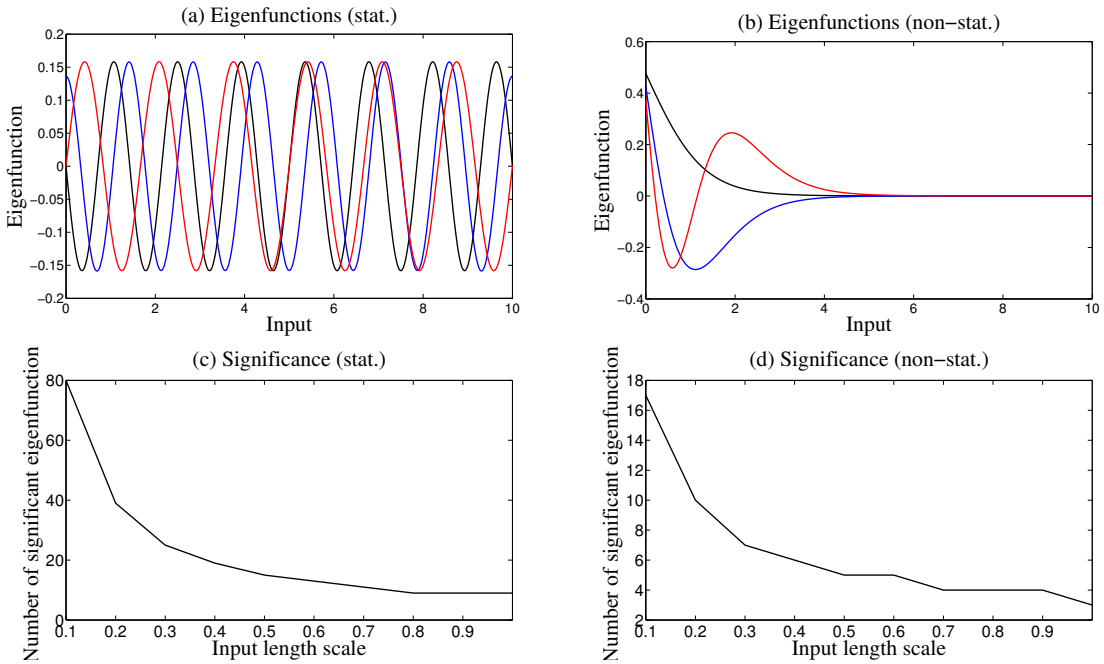


Figure 2: Example eigenfunctions for (a) stationary periodic and (b) non-stationary covariance functions, both with period 10 units. Also, the number of significant eigenfunctions for input length scales, l, for the (c) stationary periodic and (d) non-stationary covariance function.

l is the input length scale which governs the smoothness of the kernel. When the target function is periodic it is a direct function of the period *phase*, $\kappa(\tau) = |\sin(\pi\tau/D)|$ where D is the function period. Consequently, the periodic Matérn is given by Matérn$(\kappa(\tau), \nu, \sigma, 1)$. The periodic Matérn is of particular interest to us as it is used in Section 8 to model customer call centre arrival rates and in Section 9 to model the residual dynamics within home heating.

We observe that the eigenfunctions of the periodic kernel are the sinusoidal basis functions as shown in Figure 2. This basis corresponds to the Fourier basis functions for the power spectrum that can be obtained by Fourier analysis of the kernel. However, although through Fourier analysis we would be able to determine the power spectrum of the covariance function, and consequently the magnitude of the basis function, we would be unable to determine the phase of the basis function. KPCA, in contrast, is able to determine both the magnitude, and consequently phase, of the Fourier basis functions.

A key property of the KPCA approach is that the eigenfunctions are not limited to the Fourier basis and, consequently, KPCA is also able to model non-stationary periodic covariance functions efficiently, in which case the eigenfunctions, which are inferred using KPCA from the non-stationary covariance function, are anharmonic as we will now demonstrate. Figure 2(b) shows the first three most significant eigenfunctions for an exponentially

moderated periodic kernel,

$$K(t,t') = \text{Matérn}(\kappa(t-t'),\nu,\sigma,l)\exp\left(-|t|-|t'|\right) \ . \tag{32}$$

Figure 2, panels (c) and (d) show how the number of significant eigenfunctions decreases with increasing kernel smoothness for both the harmonic and anharmonic kernels above. The smoothness of the kernel is parameterised by the phase length scale, l. As above, we choose to declare an eigenfunction as significant if its eigenvalue is more than one hundredth of the maximum eigenvalue. Although this is a conservative definition of significance we can see that only a small number of basis functions are required to model these kernels.

We now compare the SSGPR, described above, and the eigenfunction approaches to modelling stationary kernels. For stationary kernels both the SSGPR and eigenfunction methods use a linear basis model with sinusoidal basis functions. The only difference between the approaches is that SSGPR assigns basis function frequencies (called *spectral points*) by sampling the kernel power spectrum. Both sine and cosine functions are used for each frequency. The KPCA infers its frequencies deterministically from the kernel and uses the basis functions with the most significant eigenvalues. Each spectral point corresponds to a Fourier basis function with known frequency with indeterminate phase. So, S spectral points produce S Fourier basis functions which has the same complexity as S Fourier basis functions in the eigenfunction approach. We compare the efficacy of both linear basis approaches when representing the squared-exponential kernel. The SSGPR was specifically developed with this kernel in mind and thus we present the fairest comparison. In order to investigate this difference and isolate the inference procedure by which the GP hyperparameters are learned from the data, the SSGPR algorithm is changed only slightly so that the actual kernel hyperparameters used correspond to the actual hyperparameters of the model which generated the training data. We also use the known generative GP hyperparameters within the eigenfunction model.

To demonstrate the superiority of the eigenfunction approach over the SSGPR approach, Figures 3 and 4 compare the SSGPR and eigenfunction representations of a squared-exponential kernel with an input scale of 10 units and output scale of 1 unit. The significant twenty two eigenfunctions were used and, equivalently, twenty two SSGPR spectral points were randomly chosen from the SE spectral density as proposed by Lázaro-Gredilla et al. (2010). Further, the eigenfunction approach used 20 evenly spaced points to construct the Gram matrix. In the case of the KPCA the corresponding covariance functions differed by no more than 9.6×10^{-5} from the actual covariance function. The SSGPR, using the same number of Fourier basis functions, deviated by as much as 0.36 (that is 36% of the prior function variance) when 22 spectral points were used. Figures 3 and 4 also show the covariance function for the SSGPR when 88 spectral points were used. In this case, the SSGPR covariance function approximation differed by as much as 0.23 (that is, 23% of the prior function variance). Clearly, the eigenfunction model is a much more accurate representation of the actual generative kernel even when using only a quarter of the number of basis functions as the SSGPR.

The error in the SSGPR representation of the covariance function can have a significant impact on the accuracy of GP inference as the SSGPR can significantly underestimate the posterior variance of the target function. To illustrate the extent of this problem, Figure 4 shows the posterior distributions of a sparsely measured function inferred using

Equations (2) and (3) and the SSGPR and eigenfunction approximations of the covariance functions. Clearly, the SSGPR variances in the top two panes are less than those calculated using the squared-exponential model (bottom right pane) and the approximate eigenfunction model (lower left pane). Furthermore, Table 1 compares the RMSE and expected log likelihood for the SSGPR and KPCA approaches over 100 functions drawn from the GP. Each function is measured every 10 units, as above, with no measurement noise. The SSGPR propensity to underestimate the posterior variance is demonstrated by a very low expected loglikelihood of -6.9×10^4 compared to 75 for the KPCA eigenfunction method. Even when the number of spectral points is increased four fold the KPCA approach is still more accurate.

In summary, the eigenfunction model is a more efficient representation than the SSGPR in that it identifies an orthogonal basis and consequently requires fewer basis functions to capture the significant features of the generative kernel. Further, as we saw earlier, the eigenfunction approach generalises to non-stationary kernels which can be represented efficiently by non-sinusoidal basis functions. Consequently, we advocate the eigenfunction approach over the SSGPR approach for generating the basis for use with LFMs.

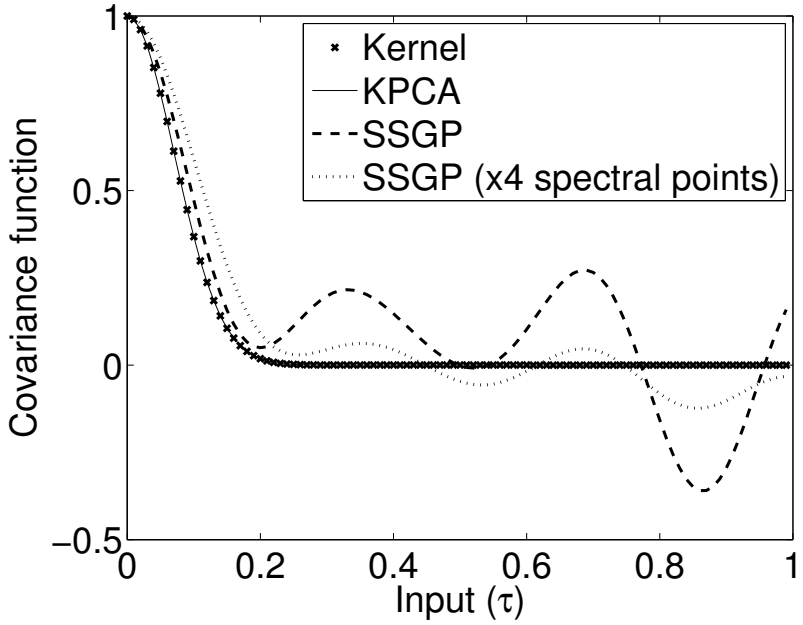


Figure 3: A comparison of SSGPR and eigenfunction approaches to modelling GP kernels via basis functions. The plots show the covariance functions corresponding to each of the eigenfunction and SSGPR models.

In the next section we extend our eigenfunction approach to quasi-periodic latent force models. This is a key contribution of our paper.

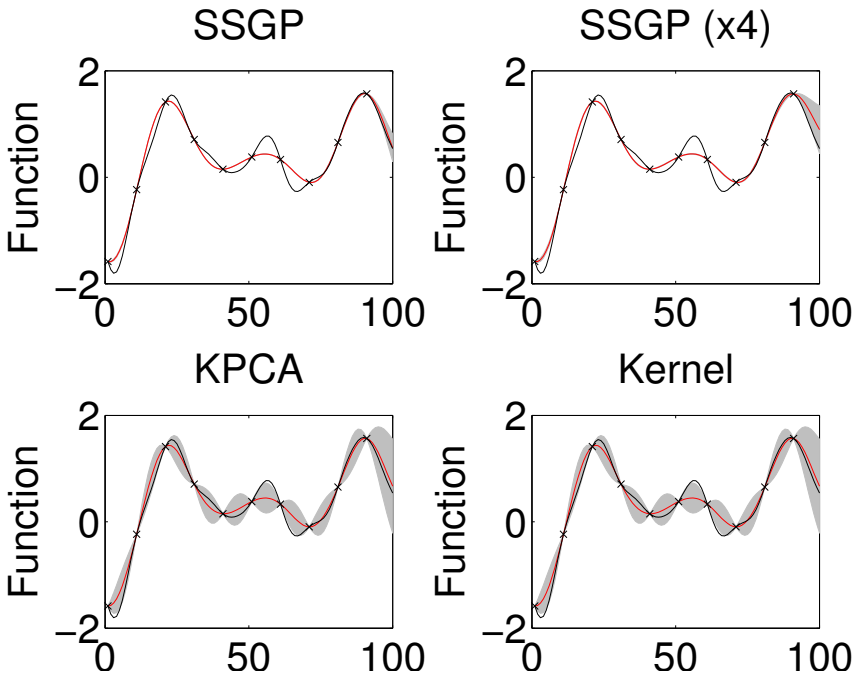


Figure 4: A comparison of SSGPR and eigenfunction approaches to modelling GP kernels via basis functions. The plots show typical example function estimates drawn using both approaches. The KPCA uses 22 basis functions and the SSGPR uses 22 spectral points and 88 spectral points respectively. The grey regions are the first standard deviation confidence regions.

Table 1: RMSE and expected log likelihood for KPCA and SSGPR with the same number of basis functions and also SSGPR with four fold increase in the number basis functions.

Method	RMSE	ELL
SSGPR	3.03 ± 0.04	$-6.9 \times 10^4 \pm 0.6 \times 10^4$
SSGPR (x4)	2.74 ± 0.03	$-1.2 \times 10^4 \pm 0.1 \times 10^4$
KPCA	$\mathbf{2.49 \pm 0.03}$	$\mathbf{75.0 \pm 0.9}$

6. REPRESENTING QUASI-PERIODIC LATENT FORCES WITH LINEAR BASIS MODELS

The eigenfunction basis model presented in the previous section assumes that the latent force is perfectly periodic. However, the force may change gradually from cycle to cycle despite the latent force kernel parameters remaining fixed. For example, the force's phase may change between cycles. In the home heating application (described in detail in Section 9),

where the residual heat within a home is modelled as a latent force, a phase shift in the residual heat profile may arise from cooking dinner at slightly different times from day to day.

When the latent force process, $u(t)$, is not perfectly periodic but exhibits some regularity from cycle to cycle it is called *quasi-periodic* and is often modelled as the product of two kernels (Rasmussen and Williams, 2006),

$$K_{\text{quasi-periodic}}(t,t') = K_{\text{quasi}}(t,t')K_{\text{periodic}}(t,t') \ , \tag{33}$$

where $K_{\text{periodic}}(t,t')$ is a periodic kernel (stationary or non-stationary) and $K_{\text{quasi}}(t,t')$ is a non-periodic kernel which reduces the inter-cycle correlations. For example, in Roberts et al. (2013), their quasi-periodic kernel is the product of a squared-exponential kernel and a periodic squared-exponential kernel,

$$K_{\text{quasi-periodic}}(t,t') = \sigma^2 \exp\left(-\frac{(t-t')^2}{l^2_{\text{quasi}}}\right) \exp\left(-\frac{\sin\left(\frac{\pi(t-t')}{D_{\text{periodic}}}\right)^2}{l^2_{\text{periodic}}}\right) \ . \tag{34}$$

We note that Equation (32) is also a quasi-periodic covariance function.

We will now demonstrate that $K_{\text{quasi-periodic}}(t,t')$ can be modelled within the state-space approach by LBMs by letting the eigenfunction weights, a as per Equation (21), change dynamically. Equation (21) can be extended to include time varying process weights, $a(t)$ (O'Hagan, 1978),

$$u(t) = \sum_j a_j(t)\phi_j(t) \ . \tag{35}$$

Thus, when $u(t)$ is generated by a quasi-periodic kernel then,

$$K_{\text{quasi-periodic}}(t,t') = E[u(t)\ u(t')] = \sum_{ij} \phi_i(t)E[a_i(t)a_j(t')]\phi_j(t') \ .$$

We assume that $a_i(t)$ is drawn from a Gaussian process,

$$a_i \sim \mathcal{GP}(0, \mu_i^\phi K_{\text{quasi}}) \ , \tag{36}$$

where μ_i^ϕ is the eigenvalue for the eigenfunction, ϕ_i, of K_{periodic} as per Equation (23). We also assume that each weight process is independent. Thus,

$$E[a_i(t)a_j(t')] = \begin{cases} \mu_i^\phi K_{\text{quasi}}(t,t') & \text{if } i = j \ , \\ 0 & \text{if } i \neq j \ . \end{cases}$$

Consequently,

$$\begin{aligned} K_{\text{quasi-periodic}}(t,t') &= \sum_i \phi_i(t)\mu_i^\phi K_{\text{quasi}}(t,t')\phi_i(t') \\ &= K_{\text{quasi}}(t,t') \sum_i \phi_i(t)\mu_i^\phi \phi_i(t') \\ &= K_{\text{quasi}}(t,t')K_{\text{periodic}}(t,t') \ . \end{aligned}$$

We see that the periodic component of the model, K_{periodic}, is represented by the basis function, ϕ, in the LBM whereas the non-periodic component, K_{quasi}, is represented via the time varying LBM coefficients, a. Note that, whereas for the resonator model, as per Equations (18) and (19), the Fourier basis functions, ϕ, are stochastic functions of time, in the eigenfunction approach, the coefficients, a, are stochastic functions of time and they reassign weight to fixed basis functions, $\phi(t)$.

In order to accommodate variant LBM coefficients in the Kalman filter we assume that each LBM coefficient is drawn from a stationary Gaussian process with covariance function, K_{quasi}, as per Equation (36). In which case, we can express the eigenfunction weight Gaussian process, $a_r(t)$, as a stochastic differential equation as per Equation (6),

$$\frac{d\mathbf{A}_r(t)}{dt} = \mathbf{F}_r\, \mathbf{A}_r(t) + \mathbf{W}_r\ \omega_r(t)\ , \tag{37}$$

where the state vector, $\mathbf{A}_r(t)$, comprises the coefficient time series and its derivatives, $\mathbf{A}_r(t) = (a_r(t)\ \frac{da_r(t)}{dt}\ , \cdots\ , \frac{d^{p_r-1}a_r(t)}{dt^{p_r-1}})^T$. Thus, as $\mathbf{A}_r$ can be expressed as a stochastic differential equation then it can be inferred using the Kalman filter as demonstrated in Hartikainen and Särkkä (2010). We can weaken the stationarity assumption and thus permit a greater choice for K_{quasi} by allowing changes in K_{quasi}'s output scale at discrete time instances called *changepoints*.

We propose three forms for K_{quasi} which are the *Continuous Quasi model* (CQM), the *Step Quasi model* (SQM) and the *Wiener-step Quasi model* (WQM). Although many other quasi-periodic forms are possible these models are chosen as they can each be represented efficiently within the Kalman filter state vector, as we will see in Section 7, whilst capturing the key qualitative properties of the data we wish to model. Specifically, the CQM models smooth, continuous deviations from cyclic behaviour over time, and, consequently, closely resembles the quasi-periodic model in Roberts et al. (2013). Alternatively, the SQM and WQM impose stationarity within a cycle but allow for function variation between cycles. We demonstrate that each can be represented in the Kalman filter via a single variable in the state-vector.

Continuous Quasi Model (CQM): This stationary model imposes changes in the cycle continuously over time t. It is equivalent to the Matérn kernel with order $\nu = 1/2$,

$$K_{\text{quasi}}^{\text{CQM}}(t, t') = \sigma_r^2 \exp\left(-\frac{|t-t'|}{l_r}\right)\ . \tag{38}$$

The input hyperparameter, l_r, is positive. As the CQM covariance function, K_{CQM}, is a first order Matérn, as per Equation (9), it can be represented as a Markov process, as per Equation (37). The process model, $\mathbf{F}_r$, and white noise spectral density, q_r, for the first order Matérn are presented in Equations (10) and (11). Reproducing this model here for completeness, if a is drawn from a GP with the quasi-periodic kernel in Equation (38), $a_r \sim \mathcal{GP}(0, K_{\text{quasi}}^{\text{CQM}})$, then,

$$\frac{da_r(t)}{dt} = F_r a_r(t) + \omega_r(t)\ ,$$

where, $\omega_r(t)$ is a white noise process with spectral density q_r and,

$$F_r = -\frac{1}{l_r} \ , \qquad\qquad q_r = \frac{2\sigma_r^2\sqrt{\pi}}{l_r\ \Gamma(0.5)} \ ,$$

and l_r and σ_r are the input and output scales, respectively, as per Equation (38). We note that, by using the CQM kernel as part of the quasi-periodic latent force covariance function, each LBM coefficient, $a_r(t)$, can be represented by a single variable in the Kalman filter state vector. In Section 7 we will demonstrate how this continuous time LTI model can be incorporated into a discrete time LFM model.

Step Quasi Model (SQM): This model can be used to decorrelate cycles at changepoints between cycles. This non-stationary model preserves the variance of the periodic function each side of the changepoint. However, the function's correlation across the changepoint is diminished. For times, t and t', with t and t' in the same cycle $K_{\text{quasi-periodic}}(t,t') = K_{\text{periodic}}(t,t')$. When times t and t' correspond to different cycles then $K_{\text{quasi-periodic}}(t,t') < K_{\text{periodic}}(t,t')$. If N consecutive cycles are labelled $C = 1, 2, \ldots, N$ and $C(t)$ denotes the cycle index for time t then,

$$K_{\text{quasi}}^{\text{SQM}}(t,t') = \sigma_r^2 \exp\left(-\frac{|C(t) - C(t')|}{l_r}\right) \ . \tag{39}$$

Again, the kernel input hyperparameter, l_r, is positive.

Wiener-step Quasi Model (WQM): Again, we assume the presence of changepoints between cycles. This non-stationary model increases the variance of the function at the changepoint. If N consecutive cycles are labelled $C = 1, \ldots, N$ then,

$$K_{\text{quasi}}^{\text{WQM}}(t,t') = \xi_0 + \min(C(t'), C(t))\xi_r \ , \tag{40}$$

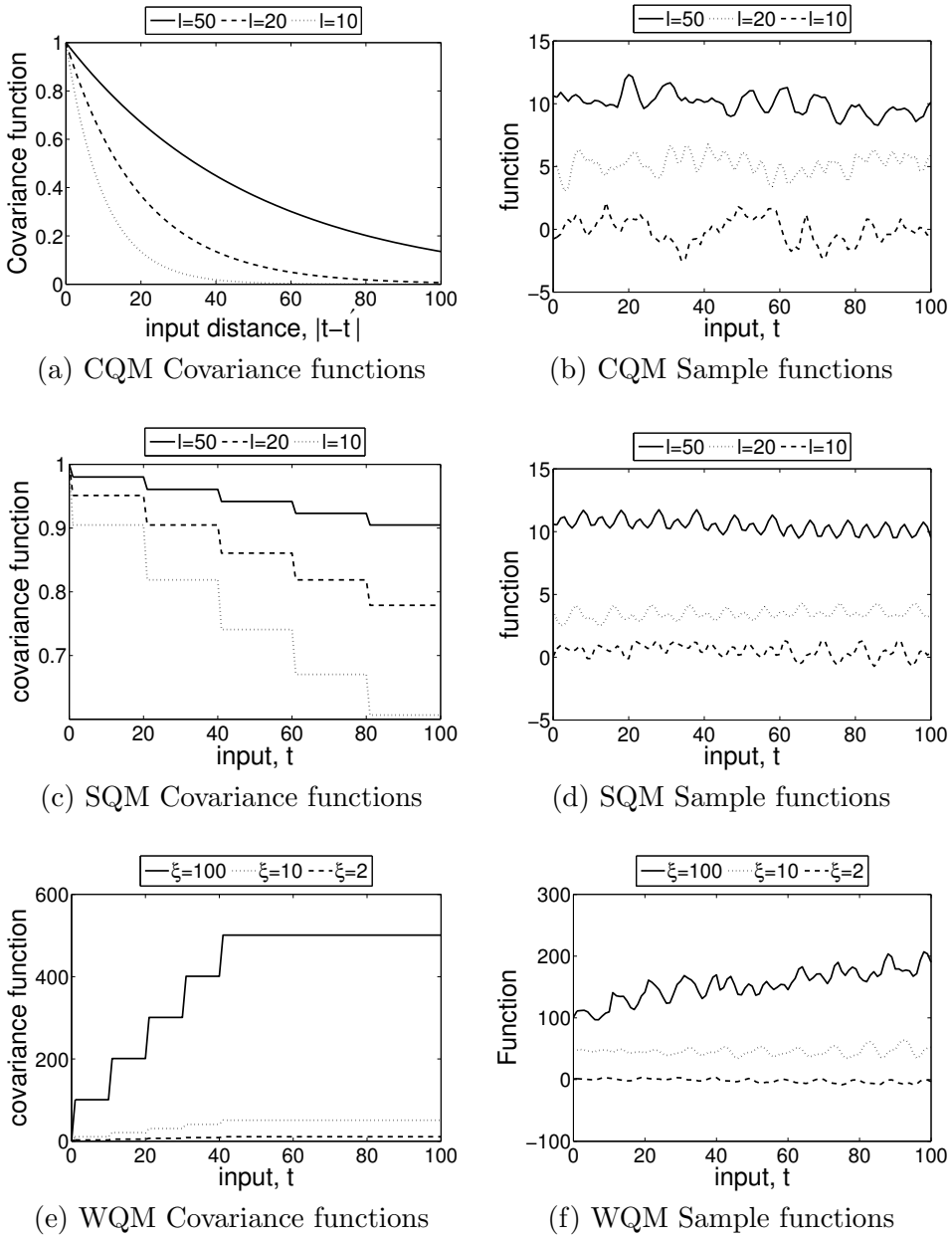
where ξ_0 and ξ_r are positive.

Example covariance functions for the three forms for K_{quasi} are shown in Figure 5. Also, sample quasi-periodic function draws are shown for each kernel. The functions are drawn from a quasi-periodic squared-exponential kernel $K_{\text{quasi-periodic}}(t,t')$ with $K_{\text{periodic}}(t,t')$ the periodic squared-exponential K_{SE}, as per Equation (16), with period $D = 10$ units, various input scales l (specified within each subfigure) and $K_{\text{quasi}}(t,t')$ set to either $K_{\text{quasi}}^{\text{CQM}}(t,t')$, $K_{\text{quasi}}^{\text{SQM}}(t,t')$ or $K_{\text{quasi}}^{\text{WQM}}(t,t')$. In the case of SQM and WQM a new cycle begins every 10 time units.

The SQM and WQM kernels can be incorporated into the discrete time Kalman filter by firstly expressing them as continuous time differential equations as per Equation (6). Suppose that either $a_r \sim \mathcal{GP}(0, K^{\text{SQM}})$ or $a_r \sim \mathcal{GP}(0, K^{\text{WQM}})$ then,

$$\frac{da_r(t)}{dt} = 0 \ ,$$

everywhere, except at changepoints. Thus, in the case of SQM and WQM the corresponding process $a_r(t)$ can be represented via a first order differential equation as per Equation (37)



(a) CQM Covariance functions (b) CQM Sample functions

(c) SQM Covariance functions (d) SQM Sample functions

(e) WQM Covariance functions (f) WQM Sample functions

Figure 5: Covariance functions (left column) for CQM, SQM and WQM. Also, sample quasi-periodic functions (right column) for CQM, SQM and WQM quasi kernels and a squared-exponential periodic kernel.

with $\mathbf{A}_r(t) = a_r(t)$, $\mathbf{\Delta}_r = 1$, $\mathbf{F}_r = 0$ and $\mathbf{W}_r = 0$. However, at a changepoint, τ, the SQM and WQM covariance functions jump in value as can be seen in Figure 5 at input distances 20 and 40, for example. The value of the process, $a_r(\tau)$, immediately after the changepoint

is related to the process, $a_r(\tau_-)$, immediately before the jump thus,

$$a_r(\tau) = G_r^* a_r(\tau_-) + \chi_r^*(\tau) \ , \tag{41}$$

where G_r^* is the process model and $\chi_r^*(\tau)$ is a Gaussian random variable, $\chi_r^*(\tau) \sim \mathcal{N}(0, Q_r^*)$. In Appendix A we demonstrate that the process model, G_r^*, and process noise variance, Q_r^*, for the SQM at the changepoint are,

$$G_{\mathrm{r,SQM}}^* = \exp\left(-\frac{1}{l_r}\right) \ , \tag{42}$$

and,

$$Q_{\mathrm{r,SQM}}^* = \sigma_r^2 \left(1 - \exp\left(-\frac{2}{l_r}\right)\right) \ , \tag{43}$$

respectively. Similarly, Appendix A also shows that the process model, G_r^*, and process noise variance, Q_r^*, for the WQM at a changepoint are,

$$G_{\mathrm{r,WQM}}^* = 1 \ , \tag{44}$$

and,

$$Q_{\mathrm{r,WQM}}^* = \xi_r \ , \tag{45}$$

respectively. The latent force variance increases at the changepoint under the WQM whereas the variance remains unchanged for the SQM. In Section 7, we demonstrate how these expressions for G^* and Q^* are incorporated within the discrete form of the Kalman filter.

The three forms for K_{quasi} will be applied to both the call centre customer queue tracking and home temperature prediction problem domains in Sections 8 and 9. In the next section we describe how we perform inference with a LFM using a state-space approach, where the state vector is augmented with periodic or quasi-periodic latent forces that are approximated using the latent force eigenfunctions.

7. RECURSIVE ESTIMATION WITH PERIODIC AND QUASI-PERIODIC LATENT FORCE MODELS

This section describes a state-space approach to inference with LFMs in some detail. We shall treat the periodic and non-periodic latent forces differently when performing inference with them. Following Hartikainen and Särkkä (2010, 2011), non-periodic forces will be modelled using the power spectrum of their corresponding covariance functions. Alternatively, the periodic latent forces will be modelled using the eigenfunctions of the corresponding periodic covariance function. The key idea in this section is to infer the LFM unknowns via the Kalman filter. The unknowns include the non-periodic forces and their derivatives, as per Equation (12), along with the coefficients of the periodic forces, as per Equation (37). The remainder of this section describes in detail how the KF state is predicted forward in time and how measurements of the system are folded into the state estimate.

We examine periodic and quasi-periodic cases separately as state-space inference with periodic latent forces uses a more compact model. For the periodic case, we assume that

the latent forces, $\mathbf{u}$, as per Equation (4), can be separated into two distinct sets, periodic forces, $\mathbf{u}_{\mathrm{p}}$, and non-periodic forces, $\mathbf{u}_{\mathrm{np}}$, so that $\mathbf{L}\mathbf{u}(t) = \mathbf{L}_{\mathrm{np}}\mathbf{u}_{\mathrm{np}}(t) + \mathbf{L}_{\mathrm{p}}\mathbf{u}_{\mathrm{p}}(t)$ as described in Section 3. Then, Equation (4) becomes,

$$\frac{d\mathbf{z}(t)}{dt} = \mathbf{F}\ \mathbf{z}(t) + \mathbf{L}_{\mathrm{np}}\mathbf{u}_{\mathrm{np}}(t) + \mathbf{L}_{\mathrm{p}}\mathbf{u}_{\mathrm{p}}(t)\ .$$

We model non-periodic latent forces and their derivatives, as per Equation (6), and periodic forces using eigenfunctions as per Equation (29). We define the augmented state vector, $\mathbf{z}_a$, as per Equation (12), and also the corresponding periodic force coefficients, $\mathbf{L}_{\mathrm{p}}^a = [\mathbf{L}_{\mathrm{p}}^T,\ \mathbf{0}^T]^T$ so that the forces $\mathbf{u}_{\mathrm{p}}$ still act on $\mathbf{z}$ within $\mathbf{z}_a$,

$$\frac{d\mathbf{z}_a(t)}{dt} = \mathbf{F}_a\ \mathbf{z}_a(t) + \mathbf{L}_a\omega_a(t) + \mathbf{L}_{\mathrm{p}}^a\mathbf{u}_{\mathrm{p}}(t)\ , \tag{46}$$

where ω_a and $\mathbf{L}_a$ are as per Equations (13) and (14).

We now introduce our eigenfunction model for the periodic latent forces into Equation (46). First, we consider periodic latent forces, introduced in Section 5, for which the corresponding LBM coefficients, $\{a\}$ in Equation (29), are constant over time. Substituting our Nyström approximation basis model for the periodic forces, as per Equation (28), into the dynamic differential model, as per Equation (46),

$$\frac{d\mathbf{z}_a(t)}{dt} = \mathbf{F}_a\ \mathbf{z}_a(t) + \mathbf{L}_a\omega_a(t) + \sum_{r=1}^{R}\sum_{j=1}^{J_r}\mathbf{L}_p^a(\cdot, r)\ \tilde{\phi}_{rj}(t)\ a_{rj}\ , \tag{47}$$

where R is the number of latent forces, J_r is the number of eigenfunctions for latent force r, a_{rj} are the eigenfunction weights and the vector $\mathbf{L}_{\mathrm{p}}^a(\cdot, r)$ is the r^{th} column of the matrix $\mathbf{L}_{\mathrm{p}}^a$ in Equation (46). The Nyström basis function, $\tilde{\phi}_{rj}$, is,

$$\tilde{\phi}_{rj}(t) = \frac{\sqrt{N_r}}{\mu_{rj}}K_r(t, S_r)\mathbf{v}_{rj}\ , \tag{48}$$

where K_r, S_r and N_r are the covariance function, the quadrature points at which the kernel is sampled for force r, as per Equation (26), and the cardinality of S_r. The μ_{rj} and $\mathbf{v}_{rj}$ are the Gram matrix eigenvalues and eigenvectors, respectively.

The differential equations (47) have the solution,

$$\mathbf{z}_a(t) = \mathbf{\Phi}(t_0, t)\mathbf{z}_a(t_0) + \mathbf{q}_a(t_0, t) + \sum_{r=1}^{R}\sum_{j=1}^{J_r} a_{rj}\mathbf{M}_{rj}(t_0, t)\ , \tag{49}$$

where, again, $\mathbf{\Phi}(t_0, t)$ denotes the matrix exponential, $\mathbf{\Phi}(t_0, t) = \exp(\mathbf{F}_a(t - t_0))$, and $\mathbf{q}_a(t_0, t) \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_a(t_0, t))$ where,

$$\mathbf{Q}_a(t_0, t) = \int_{t_0}^{t}\mathbf{\Phi}(s, t)\mathbf{L}_a\Lambda_a\mathbf{L}_a^T\mathbf{\Phi}(s, t)^T ds\ ,$$

and Λ_a, as per Equation (15), is the spectral density of the white noise processes corresponding to the non-periodic latent forces. The matrix $\mathbf{M}_{rj}(t_0, t)$ is the convolution of the state transition model, $\mathbf{\Phi}$, with each of the periodic latent force eigenfunctions,

$$\mathbf{M}_{rj}(t_0, t) = \frac{\sqrt{N_r}}{\mu_{rj}} \left[\int_{t_0}^{t} ds \ \mathbf{\Phi}(s, t) \mathbf{L}_{\mathrm{p}}^{a}(\cdot, r) K_r(s, S_r) \right] \mathbf{v}_{rj} \ .$$

For small time intervals $[t_0, \ t]$, which is the case for our applications in Sections 8 and 9, $\mathbf{M}_{rj}$ can be calculated using numerical matrix exponential integration methods. Further, we note $\mathbf{\Phi}(\mathbf{t_0}, \mathbf{t})$ is stationary and this can mitigate the need to recalculate this matrix exponential at each instance of the time series.

To accommodate the latent forces within the Kalman filter we must ensure that our discrete time dynamic model, as per Equation (49), has the appropriate form. Specifically,

$$\mathbf{X}(t) = \mathbf{G}(t_0, t)\mathbf{X}(t_0) + \omega(t_0, t) \ ,$$

where the noise process, ω, is i.i.d Gaussian and zero-mean. In order to rewrite Equation (49) into the appropriate form for Kalman filter inference we define a vector, $\mathbf{a}$, as per Equation (21), which collects together the eigenfunction weights,

$$\mathbf{a} = (a_{11}, \ldots, a_{1J_1}, a_{21}, \ldots, a_{2J_2} \ldots)^T \ ,$$

and, similarly, a matrix, $\mathbf{M}$, which collects together the convolutions, $\mathbf{M}_{rj}$,

$$\mathbf{M}(t_0, t) = (\mathbf{M}_{11}(t_0, t), \ldots, \mathbf{M}_{1J_1}(t_0, t), \mathbf{M}_{21}(t_0, t), \ldots, \mathbf{M}_{2J_2}(t_0, t) \ldots) \ .$$

We further augment the state vector to accommodate the model weights, $\mathbf{a}$, corresponding to the periodic latent forces. Let,

$$\mathbf{X}(t) = (\mathbf{z}_a^T(t), \mathbf{a}^T)^T \ , \tag{50}$$

be our augmented state vector which now accommodates the derivative auxiliary variables in $\mathbf{z}_a$ required by the non-periodic forces as per Hartikainen and Särkkä (2011) and the eigenfunction weights, $\mathbf{a}$, required by the periodic forces as per our approach. When the eigenfunction weights are constant we can rewrite Equation (49),

$$\mathbf{X}(t) = \mathbf{G}(t_0, t)\mathbf{X}(t_0) + \omega(t_0, t) \ , \tag{51}$$

where,

$$\mathbf{G}(t_0, t) = \begin{pmatrix} \mathbf{\Phi}(t_0, t) & \mathbf{M}(t_0, t) \\ \mathbf{0} & \mathbf{I} \ , \end{pmatrix} \tag{52}$$

and,

$$\omega(t_0, t) = \begin{pmatrix} \mathbf{q}_a(t_0, t) \\ \mathbf{0} \end{pmatrix} \ .$$

Thus, predictions of the Gaussian process, $\mathbf{X}$, can be inferred using the Kalman filter. Of course, the model in Equation (51) can also be incorporated within the Kalman Smoother

to perform full (that is, forward and backward) regression over $\mathbf{X}(t)$ for all time t if required (Hartikainen and Särkkä, 2010). The prediction equations for the state mean, $\bar{\mathbf{X}}(t \mid t_0)$, and covariance, $\mathbf{P}(t \mid t_0)$, at time t conditioned on measurements obtained up to time t_0, are,

$$\bar{\mathbf{X}}(t \mid t_0) = \mathbf{G}(t_0, t)\bar{\mathbf{X}}(t_0 \mid t_0) \ , \tag{53}$$

$$\mathbf{P}(t \mid t_0) = \mathbf{G}(t_0, t)\mathbf{P}(t_0 \mid t_0)\mathbf{G}(t_0, t)^T + \mathbf{Q}(t_0, t) \ , \tag{54}$$

where $\mathbf{Q}(t_0, t) \triangleq \begin{pmatrix} \mathbf{Q}_a(t_0, t) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}$.

We assume that measurements, $\mathbf{y}$, are Gaussian distributed,

$$\mathbf{y}(t) = \mathbf{H}\ \mathbf{X}(t) + \eta(t) \ , \tag{55}$$

where η is zero-mean multivariate Gaussian, $\eta \sim \mathcal{N}(\mathbf{0}, \mathbf{Z})$, where $\mathbf{Z}$ is the observation noise covariance matrix and the *measurement model*, $\mathbf{H}$, extracts the appropriate elements of the state vector. These measurements can be folded into the Kalman filter in the usual way. The update equations given measurement, $\mathbf{y}(t)$, as per equation (55), are,

$$\bar{\mathbf{X}}(t \mid t) = \bar{\mathbf{X}}(t \mid t_0) + \mathbf{K}(\mathbf{y}(t) - \mathbf{H}\bar{\mathbf{X}}(t \mid t_0)) \ , \tag{56}$$

$$\mathbf{P}(t \mid t) = (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{P}(t \mid t_0) \ , \tag{57}$$

where $\mathbf{K}$ is the Kalman gain,

$$\mathbf{K} = \mathbf{P}(t \mid t_0)\mathbf{H}^T(\mathbf{H}\mathbf{P}(t \mid t_0)\mathbf{H}^T + \mathbf{Z})^{-1} \ . \tag{58}$$

The computational complexity of the Kalman gain is cubic in the cardinality of the measurement vector, $\mathbf{y}$ (that is, not necessarily a function of the cardinality of the state). The cubic cost arises from the need to invert a covariance matrix in Equation (58). For a single output Gaussian process this covariance will be a scalar. However, for multi-output Gaussian processes, when each physical process is measured, $\mathbf{y}(t)$ will be a vector of (noisy) measurements of each process at time t. In which case, the computational complexity of the Kalman gain will be cubic in the number of measured physical processes. So, although the state vector may be augmented in order to model both physical processes and latent forces, as described above, these additions will not impact on the cost of the matrix inversion in Equation (58).

We next extend our state-space approach to accommodate quasi-periodic latent forces. For the quasi-periodic latent forces the corresponding kernel LBM coefficients, $\mathbf{a}$, are functions of time, as per Equation (35). We assume that each LBM coefficient is drawn from a Gaussian process with covariance function, K_{quasi}, as per Equation (33), and we now demonstrate how these dynamic weight processes, $\mathbf{a}(t)$, are incorporated into the Kalman filter, Equations (53) to (57).

As above, a_{rj}, corresponds to the jth eigenfunction for latent force r. However, for quasi-periodic latent forces each eigenfunction weight is variant and we assume $a_{rj}(t)$ can be written as a stochastic differential equation as proposed in Section 6,

$$\frac{d\mathbf{A}_{rj}(t)}{dt} = \mathbf{F}_{rj}\mathbf{A}_{rj}(t) + \mathbf{W}_{rj}\ \omega_{rj}(t) \ , \tag{59}$$

where the state vector, $\mathbf{A}_{rj}(t)$, comprises derivatives of the coefficient time series, $\mathbf{A}_{rj}(t) = (a_{rj}(t)\ , \frac{da_{rj}(t)}{dt}\ , \cdots\ , \frac{d^{p_{rj}-1}a_{rj}(t)}{dt^{p_{rj}-1}})^T$. We can recover the eigenfunction coefficient from $\mathbf{A}_{rj}$,

$$a_{rj}(t) = \mathbf{\Delta}_{rj}\mathbf{A}_{rj}(t)\ ,$$

where the vector $\mathbf{\Delta}_{rj} = (1, 0, \ldots, 0)$ is an indicator vector which extracts the LBM coefficient a_{rj} from $\mathbf{A}_{rj}$. Thus, the latent force, $u_r(t)$, as per Equation (35), is,

$$u_r(t) = \sum_j a_{rj}(t)\phi_{rj}(t) = \sum_j \phi_{rj}(t)\mathbf{\Delta}_{rj}\mathbf{A}_{rj}(t)\ , \tag{60}$$

where ϕ_{rj} is the jth eigenfunction for the latent force r. Substituting our Nyström approximation for the eigenfunction, $\phi(t)$ as per Equation (28), into Equation (60) we get,

$$u_r(t) = \sum_j \frac{\sqrt{N_r}}{\mu_{rj}} \left[K_r(t, S_r)\right] \mathbf{v}_{rj}\mathbf{\Delta}_{rj}\mathbf{A}_{rj}(t)\ .$$

Then, substituting u_r into the differential latent force model, Equation (47),

$$\frac{d\mathbf{z}_a(t)}{dt} = \mathbf{F}_a\ \mathbf{z}_a(t) + \mathbf{L}_a\omega_a(t) + \sum_{r=1}^{R}\sum_{j=1}^{J_r}\mathbf{m}_{rj}(t)\mathbf{A}_{rj}(t)\ , \tag{61}$$

where R is the number of latent forces, J_r is the number of eigenfunctions for latent force r, $\mathbf{L}_a$ and $\omega_a(t)$ are as per Equations (13) and (14) and the vector $\mathbf{m}_{rj}$,

$$\mathbf{m}_{rj}(t) = \frac{\sqrt{N_r}}{\mu_{rj}} \left[\mathbf{L}_{\mathrm{p}}^a(\cdot, r)K_r(t, S_r)\right] \mathbf{v}_{rj}\mathbf{\Delta}_{rj}\ , \tag{62}$$

where K_r, S_r and N_r are the covariance function, the quadrature points at which the kernel is sampled for force r, as per Equation (26), and the cardinality of S_r. The μ_{rj} and $\mathbf{v}_{rj}$ are the Gram matrix eigenvalues and eigenvectors, respectively. The vector $\mathbf{L}_{\mathrm{p}}^a(\cdot, r)$ is the r^{th} column of the matrix $\mathbf{L}_{\mathrm{p}}^a$ in Equation (46).

Now, as for the constant eigenfunction coefficient case, to exploit the Kalman filter for LFM inference with quasi-periodic latent forces we gather together all the LFM Gaussian variables, $\mathbf{z}_a$ and $\{\mathbf{A}_{rj}\}$, into a single state-vector. In so doing, we define a vector $\mathbf{A}(t)$ which collects together the eigenfunction coefficients and their derivatives,

$$\mathbf{A}(t) \triangleq (\mathbf{A}_{11}(t)^T, \mathbf{A}_{12}(t)^T, \ldots, \mathbf{A}_{21}(t)^T, \mathbf{A}_{22}(t)^T, \ldots,)^T\ , \tag{63}$$

a matrix $\mathbf{m}(t)$ which collects together the vectors $\{\mathbf{m}_{rj}\}$,

$$\mathbf{m}(t) \triangleq (\mathbf{m}_{11}(t), \mathbf{m}_{12}(t), \ldots, \mathbf{m}_{21}(t), \mathbf{m}_{22}(t), \ldots,)\ ,$$

a matrix $\mathbf{F}_{\mathbf{A}}$ which collects together the process models for all eigenfunction coefficients for all latent forces, as per Equation (59),

$$\mathbf{F}_{\mathbf{A}} \triangleq \mathrm{blockdiag}(\mathbf{F}_{11}, \mathbf{F}_{12} \ldots, \mathbf{F}_{21}, \mathbf{F}_{22}, \ldots)\ ,$$

a vector $\omega_{\mathbf{A}}$ which collects together the noise processes for the non-periodic forces, ω_a, as per Equation (61), and noise processes for all the eigenfunction coefficients as per Equation (59),

$$\omega_{\mathbf{A}} \triangleq (\omega_a^T, \omega_{11}, \omega_{12}, \ldots, \omega_{21}, \omega_{22}, \ldots)^T \ , \tag{64}$$

and the block diagonal matrix $\mathbf{L_A}$ which collects together the corresponding process noise coefficients, $\mathbf{L}_a$, as per Equation (61) and $\mathbf{W}_{rj}$ as per Equation (59),

$$\mathbf{L_A} \triangleq \text{blockdiag}(\mathbf{L}_a, \mathbf{W}_{11}, \mathbf{W}_{12} \ldots, \mathbf{W}_{21}, \mathbf{W}_{22}, \ldots) \ .$$

As per Equation (50) let,

$$\mathbf{X}(t) \triangleq (\mathbf{z}_a^T(t), \mathbf{A}^T(t))^T \ , \tag{65}$$

be our augmented state vector which now accommodates the derivative auxiliary variables required by the non-periodic forces as per Hartikainen and Särkkä (2011) and the eigenfunction weights required by the quasi-periodic forces as per our approach. Combining Equations (59) and (61),

$$\frac{d\mathbf{X}(t)}{dt} = \begin{pmatrix} \mathbf{F}_a & \mathbf{m}(t) \\ \mathbf{0} & \mathbf{F_A} \end{pmatrix} \mathbf{X}(t) + \mathbf{L_A}\omega_{\mathbf{A}}(t) \ , \tag{66}$$

where $\omega_A(t)$ is a vector of independent white noise processes. The spectral density of the ith white noise process in this vector is $[\Lambda_A]_i$ where,

$$\Lambda_A = (\Lambda_a, q_{11}, q_{12}, \ldots, q_{21,}, q_{22}, \ldots) \ .$$

The Λ_a, as per Equation (15), is the spectral density of the white noise processes corresponding to the non-periodic latent forces and q_{rj}, as per Equation (59), is the spectral density of the white noise process for the eigenfunction weight, a_{rj}.

Unfortunately, Equation (66) is an inhomogeneous SDE as $\mathbf{m}$ is a function of time. Consequently, in this form, $\mathbf{X}(t)$ cannot be folded into the Kalman filter. However, by assuming $\mathbf{m}(t)$ is approximately constant over the short time interval, $[t_0\ , t]$, and asserting $\mathbf{m}(t) \approx \mathbf{m}(t_0)$ then $\frac{d\mathbf{X}(t)}{dt}$ can be integrated into the appropriate form,

$$\mathbf{X}(t) = \mathbf{\Phi}(t_0, t)\mathbf{X}(t_0) + \mathbf{q}(t_0, t) \ , \tag{67}$$

where,

$$\mathbf{\Phi}(t_0, t) = \exp\left[\begin{pmatrix} \mathbf{F}_a & \mathbf{m}(t_0) \\ \mathbf{0} & \mathbf{F_A} \end{pmatrix} (t - t_0) \right] \ . \tag{68}$$

The process noise, $\mathbf{q}(t_0, t) \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}(t_0, t))$, where,

$$\mathbf{Q}(t_0, t) = \int_{t_0}^{t} \mathbf{\Phi}(s, t)\ \mathbf{L}_A \Lambda_A \mathbf{L}_A^T \mathbf{\Phi}(s, t)^T ds \ . \tag{69}$$

Thus, the state $\mathbf{X}(t)$ can be predicted using the Kalman filter, as per Equations (53) and (54), by defining the process model as per Equation (68),

$$\mathbf{G}(t_0, t) \quad = \quad \mathbf{\Phi}(t_0, t) \ ,$$

and the process noise covariance, $\mathbf{Q}(t_0, t)$, as per Equation (69). We note the quasi-periodic covariance functions *Step Quasi* (SQM) and *Wiener-step Quasi* (WQM), eigenfunction coefficients are perturbed, as per Equation (41), at changepoints. The discrete form of the Kalman filter can readily predict the value of each coefficient across a changepoint using the process model, G_j^*, and process noise variance, Q_j^*, for each coefficient, a_{rj}, as per Equation (41).

In general, our Kalman filter approach to LFM inference requires a process model, $\mathbf{A}_{rj}$, for each eigenfunction coefficient. Thus, the computational complexity of the prediction step of the Kalman filter which employs quasi-periodic models increases quadratically with the number of latent forces R, the number of derivatives used to represent each non-periodic latent force (N in Equation (64)) and the number of derivatives used to represent each time variant eigenfunction coefficient. Although, our approach supports any quasi-periodic covariance function, for practical applications, we recommend using the quasi-periodic covariance functions developed in Section 6 which are readily convertible to the Markovian form as per Equation (59) and for which only one variable is required to represent each time varying eigenfunction coefficient. These quasi-periodic covariance functions are the *Continuous Quasi model* (CQM), the *Step Quasi model* (SQM) and the *Wiener-step Quasi model* (WQM).

In Sections 8 and 9 we determine the efficacy of our state-space approach to LFM inference on two real world applications: i) the inference of call centre customer arrival rates and the tracking of customer queue lengths and ii) the inference of periodic residual heat dynamics within real homes and the prediction of internal temperature. We compare our approaches for the different periodic and quasi-periodic kernels developed in Section 6 on the call centre application and demonstrate the utility of incorporating periodic latent force models over non-periodic models. Then we compare our approaches to periodic and quasi-periodic LFMs to the resonator model in the thermal application.

8. MODELLING QUEUES WITH QUASI-PERIODIC ARRIVAL RATES

In this section we apply our approach to LFM inference to the dynamics of telephone queues in call centres as outlined in Section 1 with the aim of tracking the diurnal customer queue length when different agent deployment strategies are used. We use real customer arrival rate data, provided by Feigin et al. (2006), in which the customer telephone arrival rates for a loan company sales line have been collected every 5 minutes over a three month period starting from October 2001. The arrival rates during eleven consecutive Thursdays over this period are shown in Figure 6.

To model the queue dynamics as a latent force model we use the *Pointwise Stationary Fluid Flow Approximation* (PSFFA) for queues (Wang et al., 1996). The PSFFA models the *mean* queue length, L, in terms of arrival processes, ζ, using a first order differential equation,

$$\frac{dL(t)}{dt} = g(L) + \zeta(t) \tag{70}$$

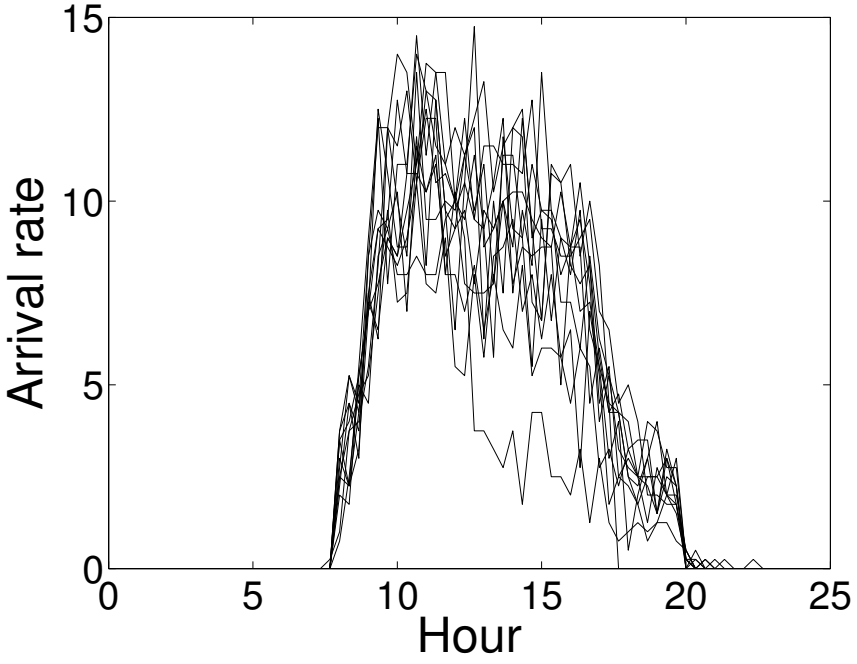


Figure 6: Customer arrival rates (per minute) for the same week day (Thursday) over a 11 week period showing the quasi-periodic nature of the data.

where, g, is a non-negative, non-linear function of the queue length, L. This model is often used to model queues in call centres where $L(t)$ is the *average* length of the queue at time t and $\zeta(t)$ is the mean *arrival rate*, that is the average rate at which customers join the queue (Wang et al., 1996). The PSFFA is a first order, non-linear differential equation. Ignoring the non-linearity of g for now, we see that Equation (70) is of the form of Equation (4). Thus, Equation (70) is an example of a LFM in which the queue length, L in Equation (70) is the target process, $\mathbf{z}$ in Equation (4) and the customer arrival rate, $\zeta(t)$ in Equation (70), is the sole latent force, $\mathbf{u}$, in Equation (4). Consequently, we apply our approach to LFM inference to the tracking of queue lengths using the PSFFA.

We consider the M/M/1 queue as it corresponds exactly to the customer arrival process, ζ, which is Poisson and the service time is arbitrarily distributed with successive service times being independent and identically distributed. [5] Service times have an exponential distribution with parameter Ω in the M/M/1 queue. A single server serves customers one at a time from the front of the queue, according to a first-come, first-served basis. When the service is complete the customer leaves the queue and the number of customers in the system reduces by one. The queue buffer is of infinite size, so there is no limit on the number of customers it can contain.

5. The term 'M/M/1' is Kendall's queue classification notation (Kendall, 1953) corresponding to a stochastic process whose state-space is the set $\{0, 1, 2, 3, ...\}$ where, in our case, the value corresponds to the number of customers in the system.

The PSFFA allows us to represent this M/M/1 system via the following differential equation for the mean queue length, L (Wang et al., 1996),

$$\frac{dL(t)}{dt} = -\Omega(t)\left(\frac{L(t)}{1+L(t)}\right) + \zeta(t) \ , \tag{71}$$

where $\Omega(t)$ is the mean queue *service rate*. We use this model to simulate the true queue length, L, using the real customer arrival rate, ζ, from data provided by Feigin et al. (2006) and realistic service rate profiles, Ω. Measurements of the instantaneous customer queue length, $y(t^*)$, are generated for times, t^*, during the day,

$$y(t^*) = L(t^*) + \epsilon(t^*) \ ,$$

where, t^*, are sufficiently spaced so that $\epsilon(t^*)$ is zero-mean, i.i.d. Gaussian.

To recover the customer arrival rate from the measured queue length we assume that the mean arrival rate, ζ, is drawn from a Gaussian process, $\zeta \sim \mathcal{GP}(0, K_{\text{arrival rate}})$, where $K_{\text{arrival rate}}$ is the arrival rate process covariance function. Note that the arrival rate can be positive or negative. Negative arrival rates correspond to customers who leave the system without being served. We choose $K_{\text{arrival rate}}$ to be either the first order Matérn kernel, a periodic first order Matérn kernel as per Equation (31) or a quasi-periodic first order Matérn kernel utilising a CQM, SQM or WQM kernel, as per Section 6.

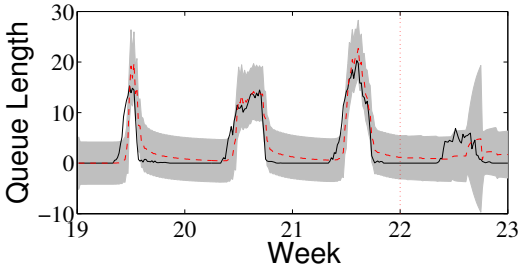
As per Equation (65), the augmented state-vector, $\mathbf{X}(t)$, is $\mathbf{X}(t) = \left(L(t), \mathbf{A}^T(t)\right)^T$ where $\mathbf{A}(t)$ are the eigenfunction weights corresponding to the periodic latent force covariance functions, as per Equation (63). Unfortunately, the transition dynamics in Equation (71) are nonlinear. However, if we assume that the mean, $\bar{L}(t_0)$, of L, conditioned on the measurements up to time t_0 is a good approximation for L over the entire, yet small, interval $[t_0,\ t]$ then we can rewrite Equation (71) as a locally linear model,

$$\frac{dL(t)}{dt} \approx -\frac{\Omega(t)}{1+\bar{L}(t_0)}\ L(t) + \zeta(t) \ , \tag{72}$$

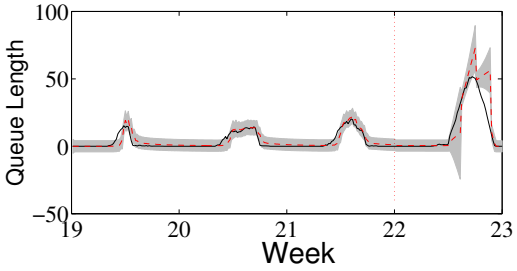
over $[t_0,\ t]$. This model then has the appropriate form for inclusion within the Kalman filter. In our experiments predictions are made over 2 minute time intervals. This time interval is chosen so that Equation (72) is a stable local approximation to the queue dynamics. We shall call this KF algorithm **LFMwith** as it contains a GP model of the arrival rate process. We use maximum likelihood to obtain the GP hyperparameters and the model parameters which are the Matérn output and input scales and the observation noise variance. The cycle period is fixed at 24 hours.

The efficacy of our customer queue model is evaluated by training the model using data over three full consecutive Thursdays and then tracking the mean queue length over the following Thursday. The queue length observations are made every 40 minutes during the training period and every three hours during the fourth day tracking phase. The longer tracking interval is specifically chosen to test the predictive power of our model with sparse observations. The efficacy of our approach to LFM inference for even longer term predictions (that is, day ahead predictions) is explored in Section 9.

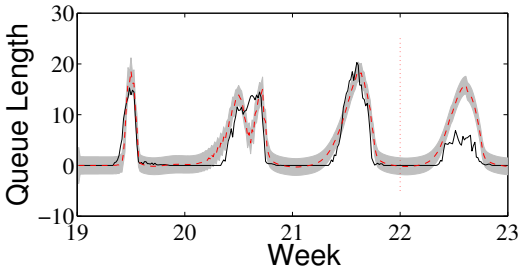
We also introduce four further algorithms to empirically demonstrate the importance of using periodic and quasi-periodic latent force models in our domain and to demonstrate



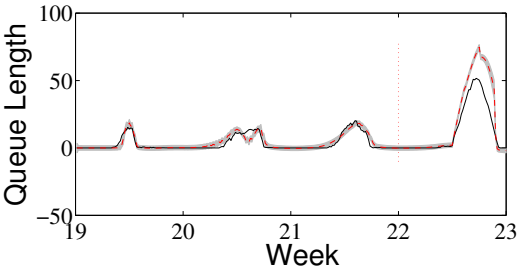
(a) **Hart**: Ω=10



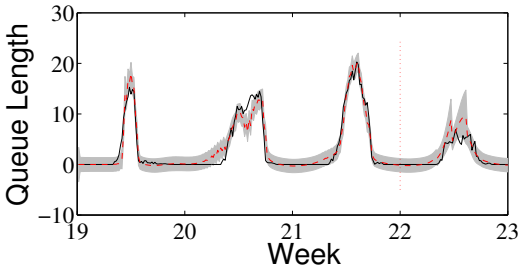
(b) **Hart**: Ω=10, 15 and then 5



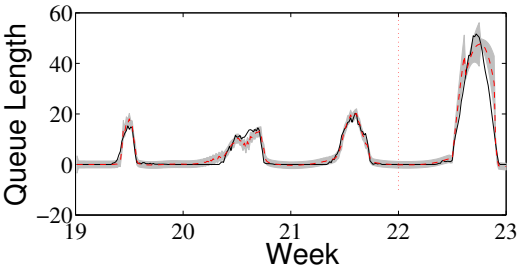
(c) **LFMwith**: Ω=10



(d) **LFMwith**: Ω=10, 15 and then 5



(e) **LFMquasi (SQM)**: Ω=10



(f) **LFMquasi (SQM)**: Ω=10, 15 and then 5

Figure 7: Call centre customer queue length over four consecutive Thursdays. The first three days of data are used to train the model. The fourth day is tracked. The 1st standard deviation confidence interval is shown (grey). The solid black line shows the ground truth. The left column of plots shows the results for a fixed service rate, $\Omega = 10$, for both training and test phases. The right column of plots shows the results for a fixed training service rate, $\Omega = 10$, and a variable test service rate of $\Omega = 15$ for the first half of the fourth day and $\Omega = 5$ for the second half.

the efficacy of our algorithm at tracking customer queue lengths. Three of these algorithms use quasi-periodic latent force models. **LFMquasi (CQM)**, **LFMquasi (SQM)** and **LFMquasi (WQM)** algorithms use the Continuous Quasi-periodic model, the Step Quasi-periodic model and the Wiener-step periodic model respectively, described in Section 6, to

model the arrival rates. These models are identical to the periodic model used in **LFMwith** except that the correlation between cycles is reduced by the quasi-periodic kernel. The most likely hyperparameters are used for the SQM, CQM and WQM kernels. The changepoints required by the SQM and WQM quasi-periodic latent force models are set to midnight for all cycles. We also implement Hartikainen's algorithm (Hartikainen and Särkkä, 2010) for sequential inference which uses the M/M/1 model described above and a non-periodic first order Matérn kernel for the customer arrival rate process (**Hart**) to demonstrate the performance of a non-periodic model.

An example run of our algorithms is shown in Figure 7 and this shows the ground truth queue lengths (in black) and first standard deviation estimates of the queue lengths using **Hart**, **LFMwith** and **LFMquasi (SQM)** latent force models. This figure shows the tracked queue length over four days. The LFM parameters are learned using the first three days of data only. The fourth day is tracked without any further learning of the LFM model parameters. The left column of plots shows the queue length estimates for a fixed service rate, $\Omega = 10$, applying to both training and tracking phases. The right column of plots shows the estimates for a fixed training service rate, $\Omega = 10$, and a variable test service rate of $\Omega = 15$ for the first half of the fourth day and $\Omega = 5$ for the second half. By testing the algorithm with variable service rates, we are able to test the efficacy of the algorithms at both reproducing the training data and at making predictions in regimes not encountered during the training period. Clearly the quasi-periodic model is the most accurate, in this case, with a RMSE of 1.9 compared to 2.3 and 4.6 for **Hart** and **LFMwith**, respectively.

To fully test the accuracy of the inferred residual model we evaluated the RMSE and expected log likelihood of the predicted average queue length for each day and for each algorithm over 11 days. Firstly, the service rate was held constant throughout at an arbitrary value of $\Omega = 10$. The results are summarised in Table 2. Clearly, the RMSE is lower for the quasi-periodic models but their expected log likelihoods are larger than the periodic model indicating the superiority of the quasi-periodic models.

Table 2: Day ahead tracking: Queue length RMSE and expected log likelihood (ELL) for periodic, quasi-periodic and non-periodic arrival rate models. Both training and test epochs have the same fixed service rate $\Omega = 10$.

Method	RMSE	ELL
LFMquasi (SQM)	4.6 ± 2.2	-146 ± 22
LFMquasi (CQM)	4.4 ± 1.3	-142 ± 15
LFMquasi (WQM)	$\mathbf{1.8 \pm 0.2}$	$\mathbf{-139 \pm 20}$
LFMwith	2.2 ± 0.4	-276 ± 60
Hart	10.6 ± 5.9	-209 ± 29

In the final experiment in this section we demonstrate the ability of our approach to make inferences in regimes where data is absent. This is a powerful and useful property of Gaussian process models. Specifically, to plan future staffing requirements the call centre manager needs to predict the impact that a novel service rate will have on future queue

lengths given predicted customer arrival rates. However, the service centre may not have utilised this staffing profile to date. In this case, for illustrative purposes, we assume that the staff profile to date has been constant during working hours with a fixed service rate, $\Omega = 10$. However, the service manager has noticed a significant queue of customers forming in the morning and then relatively few customers arriving in the afternoon. Consequently, the service manager contemplates employing a variable staffing profile and hiring more staff during the first half of the day, so that the service rate increases to $\Omega = 15$, and then retaining fewer staff in the afternoon, so that the service rate drops to $\Omega = 5$.

To determine the efficacy of our approach at predicting the impact of variable staffing profiles given only data from constant staffing profiles we repeated the experiment above with a fixed service rate, $\Omega = 10$ during training and a variable service rate during the test period. In this case, we chose a high service rate of $\Omega = 15$ for the first half of the test day and then a low rate, $\Omega = 5$ over the second half. The RMSE and expected loglikelihood are shown in Table 3. Again, the quasi-periodic models have similar efficacy and produce the most accurate estimates of the customer queue length in this case with an RMSE of 3.3 compared to 11.6 and 5.7 for **Hart** and **LFMwith**, respectively. We note that, for both experiments, the **LFMquasi (SQM)**, **LFMquasi (CQM)**, **LFMquasi (WQM)** and **LFMwith** used fewer than 28, 28, 20, 28 basis functions, respectively, to represent the arrival rate process. [6] Consequently, our LBM Kalman filter approach to LFM tracking is computationally efficient.

Table 3: Queue length RMSE and expected log likelihood (ELL) for periodic and quasi-periodic arrival rate models. Training over three days with a fixed service rate $\Omega = 10$. The test day had a service rate of $\Omega = 15$ for the first half of the day followed by $\Omega = 5$ for the remainder.

Method	RMSE	ELL
LFMquasi (SQM)	7.2 ± 1.8	-183 ± 21
LFMquasi (CQM)	15.2 ± 4.0	-205 ± 24
LFMquasi (WQM)	$\mathbf{3.3 \pm 0.6}$	$\mathbf{-152 \pm 15}$
LFMwith	5.7 ± 1.2	-725 ± 301
Hart	11.6 ± 1.1	-305 ± 49

In the next section we evaluate our approach to LFM inference for longer term predictions than those considered in the call centre application. We shall demonstrate that our approach can exploit the quasi-periodic nature of the latent force to project far forward in time an accurate estimate of the force. Consequently, we shall see that our approach is effective at performing day ahead predictions of temperatures in the home using differential thermal models and non-parametric models of the residual heat within the home.

6. The actual number of basis functions used varied between runs.

9. MODELLING THE THERMAL DYNAMICS OF HOME HEATING

In this section we apply our approach to LFM to the thermal modelling problem outlined in Section 1. We assume that the differential equation governing the thermal dynamics within a home is given by,

$$\frac{dT_{\text{int}}(t)}{dt} = \alpha(T_{\text{ext}}(t) - T_{\text{int}}(t)) + \beta E(t) + R(t) \ , \tag{73}$$

where T_{int} and T_{ext} are the internal temperature within the home and the onsite ambient external temperature respectively in °C (Bacher and Madsen, 2011; Rogers et al., 2011; Ramchurn et al., 2012). $E(t)$ represents the thermostat control output at time t ($E(t) \in \{0, 1\}$), β represents the thermal output of the heater and α is the leakage coefficient to the ambient environment. In this model $T_{\text{ext}}(t)$ and $E(t)$ are known latent forces for the LFM in Equation (73). $R(t)$ is the residual generated by latent forces which are not captured by the differential thermal model, such as heat generated by solar warming and lags in the heating system. These are completely unknown a priori but, since they are expected to exhibit periodic behaviour, a periodic Matérn kernel prior is used to model them.

We assume that $T_{\text{ext}}(t)$ at times t and t_0 can be modelled by a non-periodic GP prior Matérn($|\mathtt{t} - \mathtt{t_0}|, \mathtt{0.5}, \sigma_{\text{ext}}, \mathtt{l}_{\text{ext}}$). We choose the Matérn kernel as this imposes continuity in the function but imposes no strong assumptions about higher order derivatives. However, our approach can be applied to Matérn functions of higher smoothness if required. The state vector, $\mathbf{X}$, as per Equation (51), comprises the internal temperature, the external temperature and its derivative and eigenfunction coefficient weights, $\mathbf{A}$, for our sparse basis model of the residual as per Equation (63). We model the residual process by a periodic Matérn kernel, Matérn($|\sin(\pi\tau/\mathtt{D})|, \nu, \sigma, \mathtt{l}$) with order $\nu = 1/2$, smoothness, l, and D set to correspond to a daily period. Again, we choose the Matérn for the same reasons as above. The residual is represented via J basis functions $(\phi_1(\theta), \ldots, \phi_J(\theta))$ corresponding to Equation (28), where θ is the periodic phase as described in Section 7. The augmented state-vector, $\mathbf{X}(\mathbf{t})$, is $\mathbf{X}(t) = \left(T_{\text{int}}(t), T_{\text{ext}}(t), \frac{T_{\text{ext}}(t)}{dt}, \mathbf{A}(t)^T\right)^T$ and the continuous time dynamic model corresponding to Equation (66) is,

$$\frac{d\mathbf{X}(t)}{dt} = \begin{pmatrix} \mathbf{F}_a & \mathbf{m}(t) \\ \mathbf{0} & \mathbf{F}_{\mathbf{A}} \end{pmatrix} \mathbf{X}(t) + \mathbf{L}_{\mathbf{A}}\omega_{\mathbf{A}}(t) + \beta\mathbf{E}(t) \ ,$$

where $\mathbf{E}(t) = (E(t), 0, \ldots, 0)^T$ is the same size as $\mathbf{X}(t)$.

We will now describe the role of each term in the dynamic model. Within the transition model, $\mathbf{F}_a$ captures the temperature gradient components of Equation (73) and the Matérn driving forces for the external temperature,

$$\mathbf{F}_a = \begin{pmatrix} -\alpha & \alpha & 0 \\ 0 & 0 & 1 \\ 0 & -\rho_{\text{ext}}^2 & -2\rho_{\text{ext}} \end{pmatrix} \ ,$$

with $\rho_{\text{ext}} = 2/l_{\text{ext}}$. The derivative of the external temperature is represented in the state vector so that the Matérn latent force kernel can be encoded within the Kalman filter as summarised in Section 4 and described in detailed in Hartikainen and Särkkä (2010). We

set the order of this Matérn covariance function to $\nu = 3/2$ as the external temperature process is relatively smooth. The matrix $\mathbf{m}$ captures the residual heat contribution to the internal temperate and depends on the choice of the residual model covariance function, K, in Equation (62). Further, the matrix $\mathbf{F}_A$ models the dynamics of each periodic residual heat process and this also depends on the choice of residual model covariance function, as per Equation (59). The corresponding discrete form of the Kalman filter, as per Equation (67), is evaluated over 10 minute time intervals, $[t_0,\ t]$. This time interval is chosen to coincide with the heater on/off control cycle.

The Kalman filter is initialised with known current temperature values. The initial covariance is block diagonal with a diagonal matrix over the temperature components (including the solution to the appropriate Riccati equation for the external temperature Matérn model presented in Hartikainen and Särkkä, 2010) and a diagonal covariance over the residual model weights corresponding to the periodic Matérn residual process. The covariance for the model weights is obtained using the periodic Matérn prior and corresponding eigenfunctions as described in Section 4.

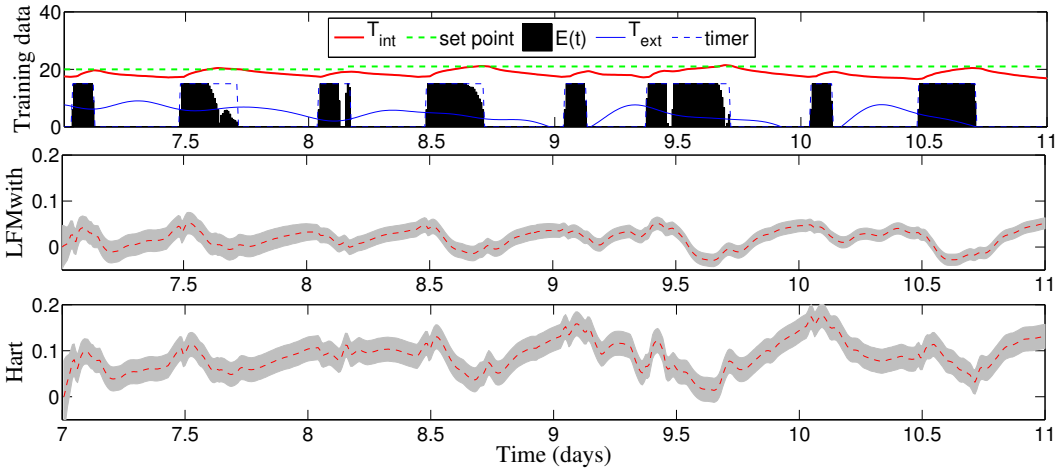


Figure 8: Internal and external temperature, thermostat set-point and heater activation for a four day training period (upper). Also shown is the residual sequentially inferred using **LFMwith** (central) and the **Hart** (lower) algorithms. The 1st standard deviation confidence interval is shown (grey).

When tracking the internal room temperature we know the state of the heater, that is, whether it is "on" (that is, $E(t) = 1$) or "off" (that is, $E(t) = 0$), at each point in time. However, when predicting the internal room temperature a full day ahead the times at which the heater will switch on or off will not be known in advance. Uncertainty in the future controller behaviour arises because the heater behaviour depends on the internal room temperature and the predicted internal room temperature will be uncertain. The heater will be on if the room temperature is below the set-point or off if above the set-point. In order

to accommodate the uncertainty in the heater switching process and, as the control output is binary, then prediction is performed using the Rao-Blackwellised Particle filter (RBPF, Doucet et al., 2000). The RBPF uses a set of particles to represent the many possible states of the system (the internal temperature and residual). The corresponding on/off control output is determined for each particle from the value of the internal temperature associated with that particle and the set point. For each particle the Kalman filter is used to predict the room temperature conditioned on the control output for that particle which is held constant for each 10 minute interval. For each 10 minute interval there are RBPF particles corresponding to the control output being "on" or "off" over that interval. Each particle also has a prior Gaussian distribution over $\mathbf{X}$ and the Kalman filter is used to predict the state $\mathbf{X}$ at the end of the current interval conditioned on the binary value of the heater for that particle. A new set of particles is then generated by taking each particle in turn, sampling the posterior of the internal temperature, conditioning the posterior on that internal temperature sample and then assigning the heater state according to the controller (set point minus the internal temperature when the heater is primed). This procedure is iterated to predict over the entire day ahead. With P particles and cardinality C of $\mathbf{X}$, the complexity of the prediction phase scales as $\mathcal{O}(C^2TP)$ over T time steps. Following nomenclature in the call centre theory application in Section 8, we shall call this RBPF algorithm **LFMwith**$^+$ as it contains a GP model of the residual heat process. However, we have added the superscript '+' to denote that LFM inference is performed by the RBPF. We use maximum likelihood to obtain the model parameters for the thermal model, $\{\alpha,\ \beta\}$, the GP hyperparameters, $\{\sigma,\ l,\ \sigma_{\text{ext}},\ l_{\text{ext}}\}$ and the observation noise variance. We note that, if the set-point process is also uncertain but a distribution over the future set-point process is known then the RBPF particles can be drawn from the on/off control distribution and the set-point distribution. We do not examine the case of uncertain set-point values in this paper.

We also implement four further algorithms to empirically demonstrate the importance of using a periodic residual model in our domain and to demonstrate the efficacy of our algorithm at predicting internal temperatures. Three of these algorithms use quasi-periodic latent force models. **LFMquasi (SQM)**$^+$, **LFMquasi (CQM)**$^+$ and **LFMquasi (WQM)**$^+$ algorithms use the step quasi-periodic model, the continuous quasi-periodic model and the Wiener quasi-periodic model, respectively, described in Section 6, to model the residual driving forces. These models are identical to the periodic model used in **LFMwith**$^+$ except that the correlation between cycles is reduced by the quasi-periodic kernel. Again, these models use the same quasi-periodic covariance functions as their counterparts in the call centre application in Section 8 and, again, we have added the superscript '+' to denote that LFM inference is performed by the RBPF. A fifth algorithm, **LFMwithout**$^+$, assumes that there is no residual heat within the home. This algorithm is an instance of **LFMwith**$^+$ with no periodic latent force basis functions in the state vector. We also implement Hartikainen's algorithm (Hartikainen and Särkkä, 2010) for sequential inference which uses the thermal model described above and a non-periodic Matérn kernel for the residual (**Hart**$^+$). To accommodate the binary thermostat controller within **Hart**$^+$ we use the RBPF, as described above, but with Hartikainen's Kalman filter formalism.

We also implement a recent version of the resonator model (Solin and Särkkä, 2013) which we call the **Resonator**$^+$. The resonators are defined via the second order differen-

tial equation, as per Equation (20), which includes a decay term with fixed frequency and decay coefficients. We chose to implement this version of the resonator model as opposed to the time varying frequency version (Särkkä et al., 2012) as this version of the resonator model is completely developed in the literature and inferring the coefficients using maximum likelihood techniques has been thoroughly tested (Solin and Särkkä, 2013). In order to undertake a fair comparison between the performance of the resonator model and our eigenfunction approaches we choose the number of resonators and eigenfunctions to be the same. Further, we infer the most likely resonator frequencies and decay coefficients using the same Nelder-Mead optimisation algorithm implemented in our eigenfunction approach. As the residual process is quasi-periodic with period, D (corresponding to one day), we initialise the resonator frequencies to be distinct and contiguous multiples of $1/D$. As with all the LFM algorithms above, day ahead predictions with the resonator model are performed by the RBPF.

We collected two data sets from two different homes in January 2012 recording the internal temperature, T_{int}, the external temperature, T_{ext}, the thermostat set point and the heater activity, E, at one minute intervals. Each data set comprises fourteen consecutive days of data. We label these data sets `data1` and `data2`. For each home four complete consecutive days of the data set are chosen to train each algorithm. We then predict the internal temperature, $T_{\text{int}}(t)$, over the next full day. With 14 days of data for each dataset, 10 full day predictions can be made for each dataset with each algorithm. Note that both data sets have thermostat set point changes that require predictions to be made in regimes in which the algorithms have not been trained.

We shall first illustrate the efficacy of the three algorithms on a single example of the training and prediction process before presenting a statistical comparison of the algorithms over the full data set. Figure 8 shows four days of training data from `data1`. The central and bottom plots show the residual over the two day period inferred by **LFMwith**$^+$ and by **Hart**$^+$. The **Hart**$^+$ plot shows that, although the residual exhibits some daily periodicity, the cycle is imperfect. However, the inferred residual for **LFMwith**$^+$ is more certain than that for **Hart**$^+$ as the periodic residual model in **LFMwith**$^+$ shares information between cycles. Consequently, the predictions drawn using **LFMwith**$^+$ are more accurate than those from **Hart**$^+$ as we will see subsequently. In addition in Figure 8 in the plot of the residual for **LFMwith**$^+$, we observe that this residual tries to compensate for the errors in the daily fall in temperature at the start of each day between times 7.0 and 7.2, 8.0 and 8.2, 9.0 and 9.2 and again between 10.0 and 10.2. We will show that these effects can have a significant impact on the day ahead prediction of the internal temperature. Although the residual model for **LFMwith**$^+$ is less certain than that for **Hart**$^+$, it captures the residual errors that arise due to using the simple thermal model in Equation (73). For instance, at the start of each day, when the heating comes on, the residual for **LFMwith**$^+$ shows a sharp spike, which represents a thermal lag in the physical process: in our homes a boiler heats up water, which, as it flows through radiators, indirectly heats up the air inside. The residual for **Hart**$^+$ however, is unable to accurately model this lag.

Figure 9 shows the day-ahead prediction of the temperature for the day immediately following the training days in Figure 8 using all seven algorithms. The **LFMquasi (CQM)**$^+$ prediction of the internal temperature has the smallest RMSE and one of the largest log

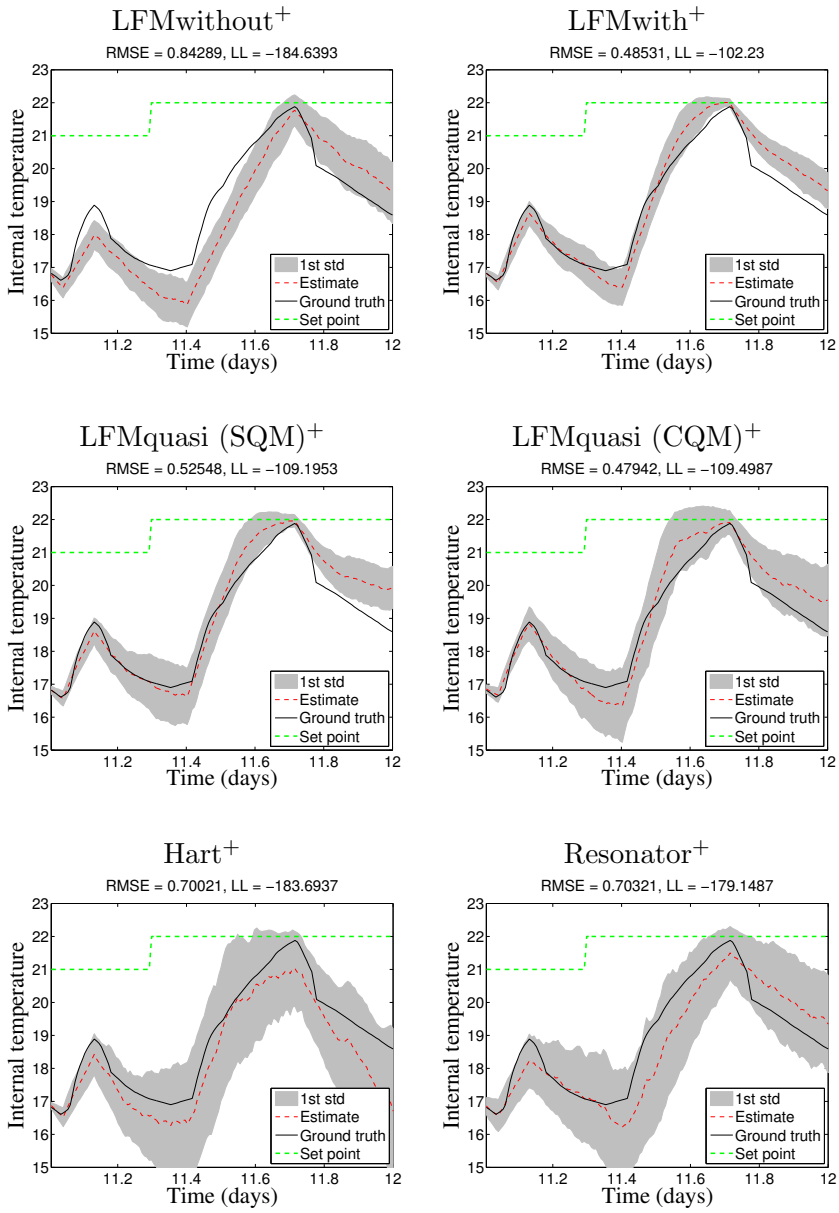


Figure 9: Internal temperature prediction compared to actual measured value using the **LFMwithout**$^+$ (top left), the **LFMwith**$^+$ (top right), the **LFMquasi (SQM)**$^+$ (middle left), the **LFMquasi (CQM)**$^+$ (middle right), the **Hart**$^+$ (bottom left) and the **Resonator**$^+$ (bottom right) algorithms. The 1st standard deviation confidence interval is shown (grey). Also shown is the thermostat set point (green).

Table 4: Internal temperature prediction RMSE of real home data `data1` over day 11 for non-periodic, quasi-periodic, periodic and no residual models.

Method	RMSE
LFMquasi $(SQM)^+$	0.53
LFMquasi $(CQM)^+$	**0.48**
$LFMwith^+$	0.49
$LFMwithout^+$	0.84
$Hart^+$	0.70
$Resonator^+$	0.70

likelihoods. This indicates that the underlying model is much more accurate than those of the other approaches. The RMSE for the example in Figure 9 is shown in Table 4.

To fully test the accuracy of the inferred residual model we evaluated the RMSE and the expected loglikelihood of the predicted temperature for each day and for each algorithm over the 10 days for both homes for which predictions were generated. The example in Figures 8 and 9 corresponds to data set `data1`, day 11. Table 5 presents the expected RMSE and the expected log likelihood of the predicted internal temperatures for each home. LFMs with periodic and quasi-periodic eigenfunction residual models have both the best RMSE and expected log likelihoods for `data1` and `data2`. The best periodic model overall with a mean RMSE of 0.52 ± 0.05 across both datasets and an expected loglikelihood of -106 ± 13 is the **LFMwith**$^+$. The **Resonator**$^+$ model has a lower consistency with an expected log likelihood of -1373 ± 1027 and it also has a higher overall RMSE at 1.15 ± 0.22. [7] The **LFMwithout**$^+$ model is weak as it is unable to accurately explain the training data without a residual model. Furthermore, although the **Hart**$^+$ approach has a very precise residual model, as shown in Figure 8, its predictions are weak since it is unaware that the residual is periodic. The non-periodic Matérn kernel, that is used by **Hart**$^+$, is unable to make accurate long term predictions of the residual since the learned input length scale for the residual is short. We note that the **LFMquasi (WQM)**$^+$ performs relatively badly on these datasets whereas the same algorithm performs well in the call centre application in Section 8. The reason for this is that the **LFMquasi (WQM)**$^+$ best models quasi-periodicity when the output scale of the residual changes between periods. Since the output scale for the heat residual does not vary from day to day then this model gives a poor fit. However, the scale of the queue length varies significantly from day to day within the call centre application and this is best modelled via the **LFMquasi (WQM)**$^+$. We note that the **LFMwith**$^+$, **LFMquasi (WQM)**$^+$, **LFMquasi (SQM)**$^+$ and **LFMquasi (CQM)**$^+$ each used fewer than 24 significant basis functions to represent the residual process. Consequently, our LBM Kalman filter approach to LFM prediction is computationally efficient.

7. The relative performance of the eigenfunction and resonator models depends on how well the model parameters are learned. Of course, changing the parameter inference mechanism could effect the performance measures reported in this paper. Given this, we endeavoured to extract the best performance from each model.

Table 5: Day ahead prediction (real home data): RMSE and expected log likelihood (ELL) for non-periodic, quasi-periodic, periodic and no residual models.

Method	data1		data2		Overall	
	RMSE	ELL	RMSE	ELL	RMSE	ELL
LFMquasi (SQM)$^+$	0.73 ± 0.22	-133 ± 24	0.59 ± 0.07	$\mathbf{-90 \pm 8}$	0.67 ± 0.13	-116 ± 15
LFMquasi (CQM)$^+$	0.85 ± 0.15	-166 ± 17	0.89 ± 0.19	-144 ± 20	0.87 ± 0.11	-157 ± 13
LFMquasi (WQM)$^+$	1.15 ± 0.14	-260 ± 17	0.94 ± 0.17	-179 ± 27	1.06 ± 0.11	-227 ± 18
LFMwith$^+$	$\mathbf{0.51 \pm 0.06}$	$\mathbf{-104 \pm 19}$	$\mathbf{0.55 \pm 0.09}$	-108 ± 19	$\mathbf{0.52 \pm 0.05}$	$\mathbf{-106 \pm 13}$
LFMwithout$^+$	0.59 ± 0.08	-156 ± 40	0.65 ± 0.11	-121 ± 21	0.61 ± 0.06	-142 ± 25
Hart$^+$	1.03 ± 0.17	-183 ± 21	0.75 ± 0.17	-130 ± 20	0.92 ± 0.12	-162 ± 16
Resonator$^+$	1.39 ± 0.32	-2122 ± 1702	0.79 ± 0.22	-250 ± 110	1.15 ± 0.22	-1373 ± 1027

We also evaluated the algorithms on the data when tracking the internal temperature over a day. We note, when tracking, the heater output is known at each time instant and, thus, it is not necessary to use the RBPF whose sole purpose is to accommodate uncertainty in the binary heater output. Thus, each LFM is now implemented through a standard Kalman filter. The LFM models were trained over four consecutive days as described above, but, in this case, the day ahead internal temperatures were filtered using measurements obtained every 100 minutes. Table 6 presents the expected RMSE and the expected log likelihood of the internal temperatures for each home.

Table 6: Tracking a day ahead: RMSE and expected log likelihood (ELL) for non-periodic, quasi-periodic, periodic and no residual models.

Method	data1		data2		Overall	
	RMSE	ELL	RMSE	ELL	RMSE	ELL
LFMquasi (SQM)	0.19 ± 0.01	85 ± 13	0.28 ± 0.04	12 ± 25	$\mathbf{0.22 \pm 0.02}$	$\mathbf{56 \pm 15}$
LFMquasi (CQM)	0.19 ± 0.02	63 ± 12	$\mathbf{0.27 \pm 0.04}$	-9 ± 26	$\mathbf{0.22 \pm 0.02}$	34 ± 15
LFMquasi (WQM)	0.26 ± 0.06	51 ± 30	0.29 ± 0.06	-23 ± 44	0.27 ± 0.04	21 ± 26
LFMwith	$\mathbf{0.18 \pm 0.02}$	$\mathbf{87 \pm 11}$	0.32 ± 0.04	-41 ± 29	0.24 ± 0.02	35 ± 21
LFMwithout	0.22 ± 0.03	48 ± 16	0.29 ± 0.04	6 ± 25	0.25 ± 0.02	31 ± 14
Hart	0.21 ± 0.02	78 ± 15	$\mathbf{0.27 \pm 0.05}$	$\mathbf{26 \pm 24}$	0.23 ± 0.02	55 ± 14
Resonator	0.82 ± 0.34	-190 ± 87	0.81 ± 0.35	-343 ± 225	0.82 ± 0.24	-251 ± 101

To determine the efficacy of the algorithms under more pronounced residual forces we simulated the heater output and, consequently, the internal temperature for residual heat drawn from a crisp quasi-periodic Matérn Gaussian process. We drew the residual process from the step-quasi model (SQM) as this model was a good representation of the real

data as demonstrated in Table 5. We then inferred the internal temperature process using Equation (73). Although we found that all three quasi-periodic models exhibited similar RMSE performance for day ahead tracking, the SQM model showed significant performance improvement over all other models when predicting a day ahead.

Example estimates for each prediction algorithm are shown in Figure 10. We re-evaluated the filter algorithms on this simulated data and the results are presented in Table 7. The **LFMquasi (SQM)$^+$** exhibits the lowest RMSE and highest loglikelihood overall with values 1.00 ± 0.16 and -190 ± 28, respectively, which isn't surprising as the alternative approaches all use incorrect models for the residual. However, specifically the **LFMquasi(SQM)$^+$** is significantly more accurate and consistent than the **Resonator$^+$** model, which has an RMSE and expected likelihood of 1.43 ± 0.19 and -279 ± 54, respectively. Consequently, despite the flexibility of the resonator model, it is unable to capture the dynamics of the SQM generated residual as it has not been informed of the prior nature of the residual and further, is unable to recover this information from the data. Clearly, encoding the appropriate prior model for the residual is critical for tracking the internal temperature accurately.

Table 7: Day ahead prediction (partially simulated home data): RMSE and expected log likelihood (ELL) for simulated non-periodic, quasi-periodic, periodic, Resonator and no residual models.

Method	data1		data2		Overall	
	RMSE	ELL	RMSE	ELL	RMSE	ELL
LFMquasi (SQM)$^+$	$\mathbf{0.75 \pm 0.24}$	$\mathbf{-161 \pm 30}$	$\mathbf{1.12 \pm 0.20}$	$\mathbf{-204 \pm 40}$	$\mathbf{1.00 \pm 0.16}$	$\mathbf{-190 \pm 28}$
LFMquasi (CQM)$^+$	1.37 ± 0.26	-288 ± 59	1.37 ± 0.23	-258 ± 48	1.37 ± 0.17	-268 ± 37
LFMquasi (WQM)$^+$	1.43 ± 0.40	-236 ± 39	1.46 ± 0.27	-241 ± 27	1.45 ± 0.21	-239 ± 21
LFMwith$^+$	1.02 ± 0.34	-338 ± 198	1.23 ± 0.25	-377 ± 116	1.16 ± 0.19	-364 ± 97
LFMwithout$^+$	1.57 ± 0.38	-376 ± 128	1.70 ± 0.36	-362 ± 101	1.66 ± 0.26	-367 ± 77
Hart$^+$	1.48 ± 0.37	-247 ± 36	1.40 ± 0.22	-294 ± 80	1.43 ± 0.19	-279 ± 54
Resonator$^+$	1.48 ± 0.37	-247 ± 36	1.40 ± 0.22	-294 ± 80	1.43 ± 0.19	-279 ± 54

We also compared the run times for each algorithm. [8] We collected the time it took to train each model on four days of data, predict an entire day ahead and then track the internal temperature over that day. For each run the resonator model and **LFMquasi (SQM)$^+$** used exactly the same number of resonators and eigenfunctions, respectively. The resonator model used between 19 and 21 resonators during the experiment. Further, the resonator model was provided with a bias term to accomodate non-zero mean residuals. Figure 11(a) shows a box plot of the single output algorithm run times. The resonator algorithm is clearly the slowest as the model inference for the resonator requires a search

8. The run times were determined using a Macbook Pro with a 2.4 GHz Intel Core i7 processor and 8 GB of memory.

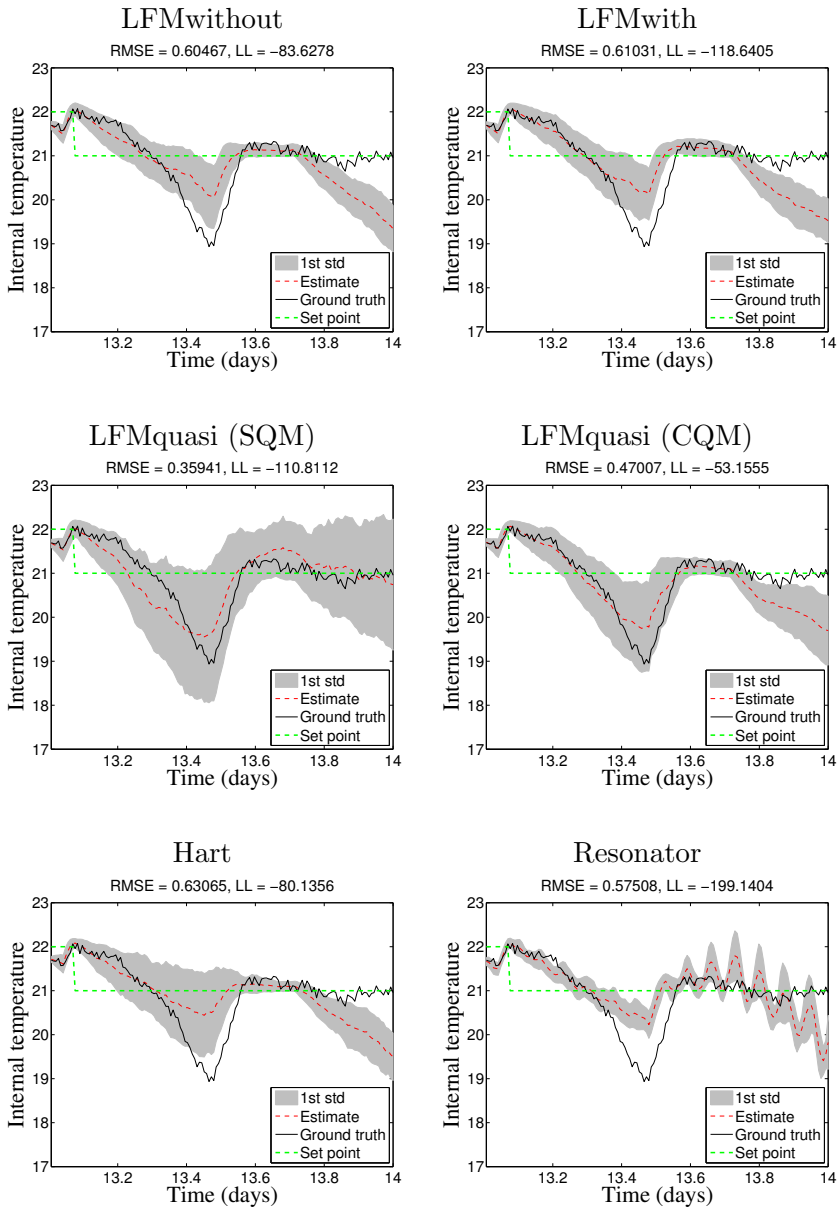


Figure 10: Predicted internal temperature compared to simulated value using the **LFMwithout** (top left), the **LFMwith** (top right), the **LFMquasi (SQM)** (middle left), the **LFMquasi (CQM)** (middle right), the **Hart** (bottom left) and the **Resonator** (bottom right) algorithms. The 1st standard deviation confidence interval is shown (grey). Also shown is the thermostat set point (green).

over a space of frequency and decay coefficients. A detailed breakdown and comparison

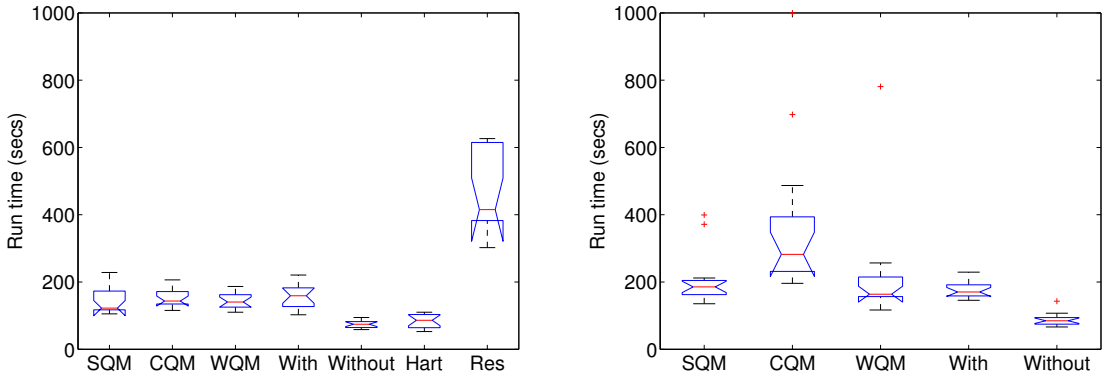
of the computational costs of the eigenfunction model and resonator model is presented in Section B.2.

Finally, we demonstrate the efficacy of our approach at modelling a multi-output system and consider an extension to the thermal model that incorporates the effect of a building's *envelope* as proposed in Bacher and Madsen (2011). The building's envelope comprises mainly the walls which act as a thermal reservoir and delay the heat transfer between the inside and the outside of the building. The multi-output model is represented by a system of coupled differential equations,

$$\frac{dT_{int}(t)}{dt} = \alpha\left(T_{env}(t) - T_{int}(t)\right) + \beta E(t) + R(t)\ , \tag{74}$$

$$\frac{dT_{env}(t)}{dt} = \Gamma\left(T_{int}(t) - T_{env}(t)\right) + \Psi\left(T_{ext}(t) - T_{env}(t)\right)\ . \tag{75}$$

Here, T_{int} and T_{env} are the internal temperature within the home and the temperature of a building's envelope, respectively. T_{env} is not directly observed, and has to be inferred from the data. The parameters in this model include: i) β, which represents the thermal output of the heater, ii) α, which regulates the convective heat transfer from the internal ambient air to the envelope, iii) Γ, which weights the convective heat transfer from the envelope to the ambient air and iv) Ψ, which represents the leakage coefficient to the ambient environment. In this model $T_{ext}(t)$ and $E(t)$ are the latent forces.



(a) Single output thermal model

(b) Multi-output thermal model

Figure 11: Empirical distribution of run times (in seconds) for each LFM algorithm for both the single output and multi-output latent force models. The time for a sample run includes the time to train the model, predict a day ahead and also track a day ahead.

To infer the internal temperature of the building using the envelope model we introduce T_{env} to the Kalman filter state vector and two further parameters, Γ and Ψ, whose most likely values are inferred from the training data. We repeated the experiments on the real

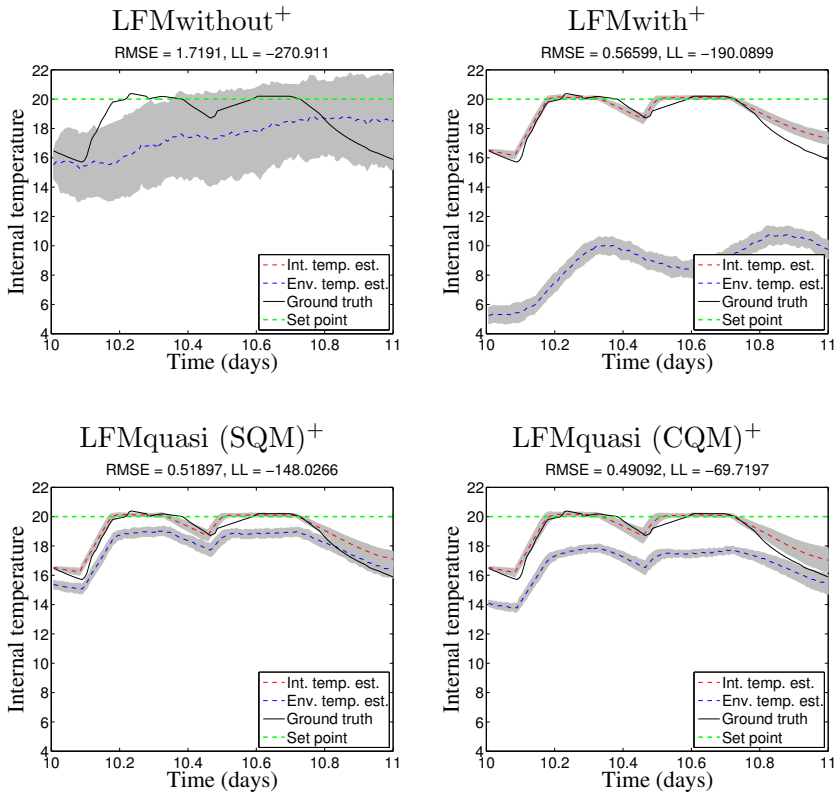


Figure 12: Internal temperature predictions compared to actual real value using the **LFMwithout**$^+$ (top left), the **LFMwith**$^+$ (top right), the **LFMquasi (SQM)**$^+$ (bottom left), the **LFMquasi (CQM)**$^+$ (bottom right) algorithms. The 1st standard deviation confidence interval is shown (grey). Also shown is the thermostat set point (green).

data above but this time using the envelope model. Figure 12 shows example day-ahead predictions of the internal and envelope temperatures for day 10 in dataset `data1`.

Table 8 presents the expected RMSE and the expected log likelihood of the predicted internal temperatures for each home. The best algorithm is the **LFMquasi (SQM)**$^+$ which uses a quasi-periodic residual model. Referring to the performance of the single-output model in Table 5 it is interesting to note that the addition of the envelope, as proposed in Bacher and Madsen (2011), improves the overall performance of all the algorithms. However, of key importance for the application of our approach to multi-output latent force models in general is the RBPF run times for this model and how they compare with the single output case. Figure 11(b) shows a box plot of the run times (the total time to train the RBPF, predict a day ahead and also track a day ahead) for the multi-output case. The run times compare favourably with the run times for the single output case despite the fact that the multi-output model requires two extra parameters. The increased computational cost for

Table 8: Predicting internal temperature a day ahead using the multi-output model: RMSE and expected log likelihood (ELL) for quasi-periodic and periodic models on real thermal data from two homes.

Method	data1		data2		Overall	
	RMSE	ELL	RMSE	ELL	RMSE	ELL
LFMquasi $(SQM)^+$	$\mathbf{0.19 \pm 0.02}$	$\mathbf{80 \pm 12}$	$\mathbf{0.27 \pm 0.04}$	7 ± 19	$\mathbf{0.22 \pm 0.02}$	$\mathbf{51 \pm 14}$
LFMquasi $(CQM)^+$	0.29 ± 0.07	12 ± 30	0.28 ± 0.04	-8 ± 18	0.29 ± 0.05	4 ± 19
LFMquasi $(WQM)^+$	0.20 ± 0.02	48 ± 27	$\mathbf{0.27 \pm 0.05}$	-1 ± 32	0.23 ± 0.02	29 ± 21
$LFMwith^+$	0.21 ± 0.02	56 ± 19	0.32 ± 0.03	-26 ± 23	0.25 ± 0.02	23 ± 18
$LFMwithout^+$	0.27 ± 0.05	24 ± 26	0.28 ± 0.05	$\mathbf{10 \pm 23}$	0.27 ± 0.04	19 ± 18

the multi-output model is due to the extra parameters in the multi-output model and this cost would be present if the standard Gaussian process inference equations, as Equations (2) and (3), were used in place of the Kalman filter. In general, the computational complexity of the Kalman filter scales quadratically with the size of the state vector and so multiple output processes can be accommodated efficiently. We note that our approach has a linear cost when conditionally independent measurements of multiple processes are incorporated. This contrasts with the standard Gaussian process inference equations which have a cubic cost in the number of processes and measurements from each process due to the need to invert a covariance matrix over all the processes.

In both the home heating application and the call centre application the eigenfunction-based models demonstrated the best performance, with improved RMSE and expected loglikelihood over non-residual models, non-periodic models and the resonator model. Further, the quasi-periodic residual models were shown to outperform perfectly periodic models on problems for which regular human behaviours, such as queuing as customers or heating homes through cooking or switching on the heating, have some influence. We noted that the WQM model had the best performance on the call centre application but the SQM exhibited the best predictive performance on the thermal modelling application. The WQM performed well on the call centre application because that application included residual forces, in this case arrival rates, which varied in amplitude from day to day. The SQM succeeded in the thermal modelling application because the residual heat profile varied slightly from day to day whilst maintaining a constant overall amplitude.

10. CONCLUSIONS

We have derived a novel and principled Bayesian approach to latent force modelling which accommodates both periodic and non-periodic forces. This approach can be incorporated within computationally efficient, iterative state-space approaches to inference. We are the first to demonstrate that eigenfunctions can be used to model periodic forces within state-space approaches to LFM inference and we offer the only principled approach to incorporating periodic covariance functions within a state-space approach to inference with LFMs. We

use the approach in Hartikainen and Särkkä (2010) for modelling non-periodic kernels and eigenfunction basis functions for modelling periodic kernels within a state-space approach to inference. We demonstrated that our eigenfunction approach out-performs the sparse spectrum Gaussian process regression (SSGPR) approach developed by Lázaro-Gredilla et al. (2010). Further, we demonstrated the close link between the eigenfunction model and the resonator model proposed by Särkkä et al. (2012). Consequently, we are the first to demonstrate how any periodic covariance function can be encoded within the resonator model using the covariance function's eigenfunctions. We are also the first to demonstrate that eigenfunctions can be represented via the resonator model within Kalman filters if required. Thus, we have proposed, in this paper, the only two approaches to date that are able to incorporate all types of Gaussian periodic model priors within a state-space approach to LFM inference. These priors include stationary periodic, non-stationary periodic and quasi-periodic covariance functions.

We have applied our approach to two applications: call centre customer queues and thermal modelling of homes. In detail, within the call centre application, customer arrival rates were modelled as driving forces through a differential model approximation of the Poisson arrival process. Both periodic and quasi-periodic models were developed to model the arrival rates of customers. The periodic models improve on the non-periodic model by as much as 83% in the root-mean-squared error. In the home heating application we modelled the thermal dynamics of homes where the physics of the energy exchange process is known but some of the heat generating processes are not known in advance. Our approach can learn the unknown heat dynamics from data and is able to accurately predict internal temperatures 24 hours ahead. Again, both periodic and quasi-periodic models were developed but, in this case, to model residual heat within the home. In this case the periodic models improve on the non-periodic model by reducing the RMSE by as much as 28%. Overall, the quasi-periodic models produced the lowest mean-squared-error and the highest expected log likelihood. Further, the eigenfunction model demonstrated improved performance over the resonator model. In the thermal application the eigenfunction models improve on the resonator by reducing the RMSE by as much as 74%.

In both the thermal modelling application and the call centre application the periodic residual models demonstrated the best performance, with improved RMSE and expected loglikelihood over non-residual models, non-periodic models and the resonator model. Further, the quasi-periodic residual models were shown to outperform perfectly periodic models in the presence of regular human behaviours, such as customer queues and heating homes through cooking or switching on the heating. We noted that the WQM model had the best performance for the call centre application but the SQM exhibited the best performance on the thermal application. The WQM performed well on the call centre application because that application included residual forces, in this case arrival rates, which varied in amplitude from day to day. The SQM succeeded in the thermal application because the residual heat profile varied slightly from day to day whilst maintaining a constant overall amplitude.

Both applications deployed state-space approaches to LFMs and both applications utilised the eigenfunction representation of the periodic latent forces acting on the system. These applications demonstrated the efficacy of our approach on both long term predictions and tracking problems. The applications demonstrated LFM inference on both linear (home heating) and non-linear (call centre) problems; on latent forces with constant output scale

(home heating) and variable output scale (call centre) and on purely Gaussian models (call centre) and models involving both Gaussian and binomial variables (thermal). We also demonstrated both single output Gaussian process and multi-output Gaussian process regression in the home heating application.

As we noted in the Appendix, the eigenfunction is the optimal RMSE basis model for any covariance function. However, our approach uses only the eigenfunctions derived from the covariance function prior. Consequently, an optimal J-dimensional model should adapt its basis functions to the data set and the eigenfunctions of the posterior covariance function should be used. We believe it is possible to extend our approach to accommodate adaptable eigenfunctions and this will be the focus of further work. Further, the Kalman formalism expressed in this paper lends itself immediately to control problems and we intend to investigate our approach to LFM inference within model-based predictive control. This research will be of particular value to domains in which some physical knowledge of the process is known (and expressible via differential equations) and nonparametric models can be used to express the latent forces. We will explore the relative merits of expressing control problems directly via the Gaussian process prior as in, for example, Ažman and Kocijan (2008), and via the Markovian formalism advocated in this paper.

ACKNOWLEDGEMENTS

This work was funded in the UK by the EPSRC ORCHID programme grant (EP/I011587/1) and the EPSRC 'Intelligent Agents for Home Energy Management' project (EP/I000143/1), and in Kingdom of Saudi Arabia by the Deanship of Scientific Research (DSR), King Abdulaziz University, Jeddah (9-15-1432-HiCi).

References

M. Alvarez, J. Peters, B. Schoelkopf, and N. Lawrence. Switched Latent Force Models for Movement Segmentation. In J. Lafferty, C. K. I. Williams, J. Shawe-Taylor, R. S. Zemel, and A. Culotta, editors, *Advances in Neural Information Processing Systems 23*, pages 55–63, 2010.

M. A. Alvarez and N. D. Lawrence. Sparse Convolved Gaussian Processes for Multi-Output Regression. In *Advances in Neural Information Processing Systems 21*, pages 57–64, 2008.

M. A. Alvarez and N. D. Lawrence. Computationally Efficient Convolved Multiple Output Gaussian Processes. *Journal of Machine Learning Research*, 12:1459–1500, 2011.

M. A. Alvarez, D. Luengo, and N. D. Lawrence. Latent force models. *Journal of Machine Learning Research - Proceedings Track*, 5:9–16, 2009.

M. A. Alvarez, D. Luengo, and N. D. Lawrence. Linear Latent Force Models Using Gaussian Processes. Technical report, Department of Computer Science, University of Manchester, UK, 2011. `http://arxiv.org/abs/1107.2699`.

G. B. Arfken, H. J. Weber, and F. E. Harris. *Mathematical Methods for Physicists*. Academic press, 2005.

K. Ažman and J. Kocijan. Non-Linear Model Predictive Control for Models with Local Information and Uncertainties. *Transactions of the Institute of Measurement and Control*, 30(5):371–396, 2008.

P. Bacher and H. Madsen. Identifying Suitable Models for the Heat Dynamics of Buildings. *Energy and Buildings*, 43(7):1511–1522, 2011.

Y. Bengio, O. Delalleau, N. Le Roux, J.-F. Paiement, P. Vincent, and M. Ouimet. Learning Eigenfunctions Link Spectral Embedding and Kernel PCA. *Neural Computation*, 16: 2197–2219, 2004.

C. M. Bishop. *Neural Networks for Pattern Recognition*. Oxford University Press, 1999. ISBN 0198538642.

DECC. Smarter Grids: The Opportunity. Technical report, Department of Energy and Climate Change (DECC), 2009. `http://www.decc.gov.uk`.

A. Doucet, J. F. G. de Freitas, K. Murphy, and S. Russel. Rao-Blackwellized Particle Filtering for Dynamic Bayesian Networks. In *Proc of the Conference on Uncertainty in Artificial Intelligence (UAI '00)*, pages 176—183, 2000.

P. Drineas and M. W. Mahoney. On the Nyström Method for Approximating a Gram Matrix for Improved Kernel-Based Learning. *Journal of Machine Learning Research*, 6: 2153–2175, 2005.

P. D. Feigin, A. Mandelbaum, S. Zeltyn, V. Trofimov, E. Ishay, P. Khudiakov, and E. Nadjharov. DataMocca: Data MOdel for Call Center Analysis, The Call Center of "US Bank". http://ie.technion.ac.il/Labs/Serveng/files/The_Call_Center_of_US_Bank.pdf, 2006.

J. Hartikainen and S Särkkä. Kalman Filtering and Smoothing Solutions to Temporal Gaussian Process Regression Models. In *Proc of IEEE International Workshop on Machine Learning for Signal Processing (MLSP)*, pages 379–384, 2010.

J. Hartikainen and S. Särkkä. Sequential Inference for Latent Force Models. In *Proc of the Twenty-Seventh Conference Annual Conference on Uncertainty in Artificial Intelligence (UAI-11)*, pages 311–318, 2011.

J. Hartikainen, M. Seppanen, and S. Särkkä. State-Space Inference for Non-Linear Latent Force Models with Application to Satellite Orbit Prediction. In *Proceedings of the 29th International Conference on Machine Learning (ICML '12), Edinburgh, Scotland*, 2012.

S. Ji and R. Zhou. Simulation and Bottleneck Analysis of Container Port Handling Resources Based on Bounding Theory. In *International Conference on Logistics Systems and Intelligent Management (ICLSIM '10)*, volume 2, pages 705–708, 2010.

R. E. Kalman et al. A New Approach to Linear Filtering and Prediction Problems. *Journal of basic Engineering*, 82(1):35–45, 1960.

D. G. Kendall. Stochastic Processes Occurring in the Theory of Queues and their Analysis by the Method of the Imbedded Markov Chain. *The Annals of Mathematical Statistics*, pages 338–354, 1953.

M. Lázaro-Gredilla, J. Quiñonero-Candela, C. E. Rasmussen, and A. R. Figueiras-Vidal. Sparse Spectrum Gaussian Process Regression. *Journal of Machine Learning Research*, 11:1865–1881, 2010.

M. Loéve. *Probability Theory; Foundations, Random Sequences*. New York: D. Van Nostrand Company, 1955.

D. MacKay. *Sustainable Energy: Without the Hot Air*. UIT, Cambridge, 2009.

J. Mercer. Functions of Positive and Negative Type, and their Connection with the Theory of Integral Equations. *Philosophical Transactions of the Royal Society of London. Series A*, 209:415–446, 1909.

A. O'Hagan. Curve Fitting and Optimal Design for Prediction. *Journal of the Royal Statistical Society*, 40:1–42, 1978.

E. Parzen. On Estimation of a Probability Density Function and Mode. *Annals of Mathematical Statistics*, 33:1065—1076, 1962.

J. Quiñonero Candela and C. E. Rasmussen. A Unifying View of Sparse Approximate Gaussian Process Regression. *Journal of Machine Learning Research*, 6:1939–1959, 2005.

S. Ramchurn, P. Vytelingum, A. Rogers, and N. R. Jennings. Putting the "Smarts" into the Smart Grid: A Grand Challenge for Artificial Intelligence. *Communications of the ACM*, 55(4):86–97, 2012.

C. E. Rasmussen and C. K. I. Williams. *Gaussian Processes for Machine Learning*. The MIT Press, 2006.

S. Roberts, M. Osborne, M. Ebden, S. Reece, N. Gibson, and S. Aigrain. Gaussian Processes for Time Series Modelling. *Phil. Trans. R. Soc. A*, 371(1984 20110550), 13 February 2013.

A. Rogers, S. Maleki, S. Ghosh, and N. R. Jennings. Adaptive Home Heating Control Through Gaussian Process Prediction and Mathematical Programming. In *Second International Workshop on Agent Technology for Energy Systems (ATES 2011)*, pages 71–78, 2011.

S. Särkkä, A. Solin, A. Nummenmaa, A. Vehtari, T. Auranen, S. Vanni, and F.-H. Lin. Dynamic Retrospective Filtering of Physiological Noise in BOLD FMRI: DRIFTER. *NeuroImage*, 2012.

B. Schölkopf and K. R. Müller. Nonlinear Component Analysis as a Kernel Eigenvalue Problem. *Neural Computation*, 10:1299—1319, 1998.

J. Shawe-Taylor, C. K. I. Williams, N. Cristianini, and J. Kandola. On the Eigenspectrum of the Gram Matrix and the Generalization Error of Lernel-PCA. *Information Theory, IEEE Transactions on*, 51(7):2510–2522, 2005.

M. Sims, J. Kurose, and V. Lesser. Streaming versus Batch Processing of Sensor Data in a Hazardous Weather Detection System. In *Proceedings of Second Annual IEEE Communications Society Conference on Sensor and Ad Hoc Communications and Networks (SECON 2005)*, pages 185–196, September 2005.

E. Snelson and Z. Ghahramani. Sparse Gaussian Processes Using Pseudo-Inputs. In *Advances in Neural Information Processing Systems 18*, pages 1257–1264, 2006.

A. Solin and S. Särkkä. Infinite-Dimensional Bayesian Filtering for Detection of Quasi-Periodic Phenomena in Spatio-Temporal Data. Technical report, Department of Biochemical Engineering and Computational Science, Aalto University, 2013.

M. Tipping. Sparse Bayesian Learning and the Relevance Vector Machine. *Journal of Machine Learning Research*, 1:211–244, 2001.

W.-P. Wang, D. Tipper, and S. Banerjee. A Simple Approximation for Modeling Nonstationary Queues. In *Proceedings of the Fifteenth Annual Joint Conference of the IEEE Computer and Communications Societies Conference on The Conference on Computer Communications - Volume 1*, pages 255–262, 1996.

C. Williams and M. Seeger. Using the Nyström Method to Speed Up Kernel Machines. In *Advances in Neural Information Processing Systems 13*, pages 682–688. MIT Press, 2001.

Appendix A. DISCRETE JUMP MARKOV PROCESSES FOR NON-STATIONARY COVARIANCE FUNCTIONS

We describe how the Step Quasi model (SQM) and the Wiener-step Quasi model (WQM) can be incorporated within the discrete time Kalman filter.

Suppose that either $a \sim \mathcal{GP}(0, K^{\mathrm{SQM}})$ or $a \sim \mathcal{GP}(0, K^{\mathrm{WQM}})$, where K^{SQM} and K^{WQM} are the covariance functions for the SQM and WQM, respectively. Consider a changepoint, τ and some earlier time τ_- close to τ such that $\tau > \tau_-$. We assume that,

$$a(\tau) = Ga(\tau_-) + \chi(\tau) \tag{76}$$

where G is the Kalman filter process model and $\chi(\tau) \sim \mathcal{N}(0, Q)$. We will now see that G and Q can be expressed in terms of the kernels, K^{SQM} and K^{WQM} at the changepoint τ. Recall $E[a(t)] = E[\chi(t)] = 0$, $E[a(t)\chi(t)] = 0$ and $K(t, t') = E[a(t)a(t')]$ for all t and t'. Thus, by squaring both sides of Equation (76) and then taking the expectation we get the variance, $K(\tau, \tau)$, of $a(\tau)$,

$$K(\tau, \tau) = GK(\tau_-, \tau_-)G + Q \ . \tag{77}$$

Also, multiplying Equation (76) throughout by $a(\tau_-)$ before taking the expectation gives the covariance between $a(\tau)$ and $a(\tau_-)$,

$$K(\tau, \tau_-) = GK(\tau_-, \tau_-) \ . \tag{78}$$

Specifically, from Equation (39), the SQM variance, $K^{\mathrm{SQM}}(\tau, \tau) = \sigma^2$ and $K^{\mathrm{SQM}}(\tau, \tau_-) = \sigma^2 \exp(-1/l)$ as $C(\tau) - C(\tau_-) = 1$ in Equation (39) across a single changepoint. Thus,

using Equations (77) and (78), the process model, G, and process noise variance, Q, for the SQM are,

$$G_{\text{SQM}} = \exp\left(-\frac{1}{l}\right) , \qquad Q_{\text{SQM}} = \sigma^2\left(1 - \exp\left(-\frac{2}{l}\right)\right) .$$

Also, by Equation (40) the WQM variances, $K^{\text{WQM}}(\tau,\tau) = \xi_0 + C(\tau)\xi$, $K^{\text{WQM}}(\tau_-,\tau_-) = \xi_0 + C(\tau_-)\xi$ and covariance, $K^{\text{WQM}}(\tau,\tau_-) = \xi_0 + C(\tau_-)\xi$. Thus, by Equations (77) and (78), the process model, G, and process noise variance, Q, for the WQM are,

$$G_{\text{WQM}} = 1 , \qquad Q_{\text{WQM}} = \xi$$

as $C(\tau) - C(\tau_-) = 1$.

Appendix B. COMPARISON OF EIGENFUNCTION AND RESONATOR MODELS

In this section we assert that the eigenfunction basis model advocated in this paper is optimal in that it minimises the mean squared error for all possible J-dimensional basis models and thus establish the eigenfunction approach as the preferred approach. We shall then develop the theoretical link between the eigenfunction basis model and the *resonator model* (Särkkä et al., 2012; Hartikainen et al., 2012; Solin and Särkkä, 2013) which is the most significant alternative approach to modelling periodic forces in LFMs. Consequently, we will demonstrate that the resonator model parameters can be chosen so that the resonator basis is equivalent to the eigenfunction basis. As a corollary we propose a novel mechanism for encoding periodic covariance function priors in the resonator model.

B.1 Establishing the Link Between the Resonator Basis and Eigenfunctions

A *J-dimensional linear model* is a linear combination of J basis functions. Both the eigenfunction model, as per Equation (25), and resonator model, as per Equation (18), are J-dimensional linear models. The eigenfunction model, as per Equation (25), is a linear combination of orthonormal basis functions, ϕ_j, whereas the resonator model is a linear combination of resonators, ψ_j, which are not necessarily orthogonal.

Let g be some function drawn from a Gaussian process with covariance function K. Then the Karhunen-Loéve expansion theorem (Loéve, 1955) states that the eigenfunction basis is the orthonormal basis that minimises the total mean squared error between the J-dimensional model and g. Further, any non-orthonormal basis with cardinality, α, can be converted to an orthonormal basis with cardinality, υ, such that $\upsilon \leq \alpha$, by Gram-Schmidt orthogonalisation (Arfken et al., 2005) and renormalisation. Thus, we can establish immediately that the eigenfunction basis is the optimal mean squared basis for all J-dimensional linear models. [9]

9. In this paper, we use a static basis chosen from the prior covariance function. However, the eigenfunctions are dependent on the covariance function and consequently, an optimal J-dimensional model should adapt its basis when evidence is integrated with the prior. We believe it is possible to extend our approach to accommodate adaptable eigenfunctions and this will be the focus of a further paper. We note that the resonator model can also adapt to the evidence provided the frequency process in Equation (19) adapts with the data.

In the remainder of this section we determine the conditions under which each version of the resonator model, as per Equations (19) and (20), is equivalent to the optimal eigenfunction model.

B.1.1 Perfectly Periodic and Stationary Covariance Functions

A *perfectly periodic stationary process* $g \sim \mathcal{GP}(b, K)$ with period D satisfies, $g(t + nD) = g(t)$ for all $t \in \mathbb{R}$ and $n \in \mathcal{N}$. Such functions (for example, the squared-exponential in Equation (16)) are generated from Gaussian processes with covariance functions of the form $K(t, t') = h(t - t')$ for some function h.

Bochner's theorem (see, for example, Rasmussen and Williams, 2006) states that the eigenfunctions of a stationary kernel are the Fourier basis functions. Thus, the optimal J-dimensional linear model for a stationary Gaussian process is a linear combination of Fourier basis functions. Both resonator models, in Equations (19) and (20), can model Fourier basis functions exactly by asserting $\omega_j(t) = 0$ for all time t and all resonators, j, in Equation (19), and assigning a constant resonator frequency, f, in the original resonator model, as per Equation (19), or removing the decay term by setting $B_j = 0$ in the later model, as per Equation (20), and assigning $A_j = -(2\pi f_j)^2$,

$$\frac{d^2\psi_j(t)}{dt^2} = -(2\pi f_j)^2\psi_j(t) \ .$$

Thus, the optimal J-dimensional linear model for the stationary kernel case is an instance of both resonator models.

B.1.2 Perfectly Periodic and Non-stationary Covariance Functions

A *perfectly periodic non-stationary process* $g \sim \mathcal{GP}(b, K)$ with period D satisfies, $g(t+nD) = g(t)$ for all $t \in \mathbb{R}$ and $n \in \mathcal{N}$. Such processes (for example, Equation (32)) are Gaussian processes with covariance functions of the form $K(t, t') = h(t, t')$ where $h(t, t') \neq h(t - t')$. Note that since the latent force g is perfectly periodic then the resonator cannot be stochastic (that is, $\omega_j(t) = 0$ for all time t and resonator, j, in Equation (19)).

We demonstrate that the eigenfunctions for non-stationary covariance functions can be represented by the resonator model using the time varying frequency model, as per Equation (20), provided that the eigenfunction is second order differentiable. We note that the eigenfunction linear basis model, as per Equation (29), and the resonator model, as per Equation (18), are equivalent if,

$$\psi_j(t) = a_j\phi_j(t) \ , \tag{79}$$

for eigenfunction, ϕ_j, resonator, ψ_j, times, t, and some positive coefficient, a_j, as per Equation (18). Substituting Equation (79) into Equation (19), asserting $\omega_j(t) = 0$ (as above) and rearranging,

$$(2\pi f_j(t))^2 = -\frac{1}{\phi_j(t)}\frac{d^2\phi_j(t)}{dt^2} \ . \tag{80}$$

Thus, any perfectly periodic covariance function can be encoded within the resonator model by defining the frequency process, $f_j(t)$, in terms of the covariance function eigenfunctions,

$\phi_j(t)$. Furthermore, we can also represent eigenfunctions via the resonator model within Kalman filters if required. In practise, the Nyström approximation, $\tilde{\phi}$, for the eigenfunction basis is used in place of ϕ in Equation (80) to calculate the frequency process for the resonator model. [10]

To illustrate the link between the eigenfunction and corresponding resonator models for perfectly periodic covariance functions we derive the frequency process, $f(t)$, for a variation of the non-stationary covariance function in Equation (32) with a low smoothness, $\nu = 3/2$,

$$K(t,t') = \text{Matérn}(\kappa(t-t'),\nu,\sigma,l)\exp(-\alpha(\kappa(t)^2+\kappa(t')^2)) \ , \tag{81}$$

where $\kappa(\tau) = |\sin(\pi\tau/D)|$, D is the covariance function period and $\alpha > 0$ is the decay rate. This covariance function differs from Equation (32) in two crucial respects. Firstly, it is now perfectly periodic with period D and secondly, it is second order differentiable everywhere, as required by Equation (80). For $\nu = 3/2$ the Matérn simplifies,

$$\text{Matérn}(\kappa(\tau),3/2,\sigma,l) = \sigma^2(1+\sqrt{3}\kappa(\tau)/l)\exp(-\sqrt{3}\kappa(\tau)/l) \ .$$

Using the Nyström approximation, as per Equation (28),

$$\frac{d^2\tilde{\phi}_i(t)}{dt^2} = \frac{\sqrt{N}}{\mu_i}\frac{d^2K(t,S)}{dt^2}\mathbf{v}_i \ . \tag{82}$$

After some algebra,

$$\begin{aligned}
\frac{d^2K(t,t')}{dt^2} &= \frac{d^2\text{Matérn}(\kappa(\tau),3/2,\sigma,l)}{d\tau^2}\exp(-\alpha\kappa(t)^2)\exp(-\alpha\kappa(t')^2) \\
&\quad +\text{Matérn}(\kappa(\tau),3/2,\sigma,l)\frac{d^2\exp(-\alpha\kappa(t)^2)}{dt^2}\exp(-\alpha\kappa(t')^2) \\
&\quad +2\frac{d\text{Matérn}(\kappa(\tau),3/2,\sigma,l)}{d\tau}\frac{d\exp(-\alpha\kappa(t)^2)}{dt}\exp(-\alpha\kappa(t')^2) \ ,
\end{aligned}$$

where $\tau = t - t'$ and,

$$\begin{aligned}
\frac{d\text{Matérn}(\kappa(\tau),3/2,\sigma,l)}{d\tau} &= -\frac{3\pi\sigma^2}{2Dl^2}\sin\left(\frac{2\pi}{D}\tau\right)\exp\left(-\frac{\sqrt{3}}{l}\kappa(\tau)\right) \ , \\
\frac{d\exp(-\alpha\kappa(t)^2)}{dt} &= -\frac{\pi\alpha}{D}\sin\left(\frac{2\pi}{D}t\right)\exp(-\alpha\kappa(t)^2) \ , \\
\frac{d^2\text{Matérn}(\kappa(\tau),3/2,\sigma,l)}{d\tau^2} &= \frac{3\pi^2\sigma^2}{D^2l^3}\exp\left(-\frac{\sqrt{3}}{l}\kappa(\tau)\right)\left(\sqrt{3}\kappa(\tau)(1-\kappa(\tau)^2)-l(1-2\kappa(\tau)^2)\right) \ , \\
\frac{d^2\exp(-\alpha\kappa(t)^2)}{dt^2} &= -\frac{2\pi^2\alpha}{D^2}\left(\frac{\alpha}{2}\left([1-2\kappa(t)^2]^2-1\right)+1-2\kappa(t)^2\right)\exp(-\alpha\kappa(t)^2) \ .
\end{aligned}$$

10. We note by Equation (80) the resonator can become unstable close to $\tilde{\phi} = 0$. This problem is easily solved by initially adding some offset, Δ, to $\tilde{\phi}$ for some suitably large Δ before calculating the frequency process $f(t)$. Consequently, when $f(t)$ is used in the resonator model, as per Equation (19) the corresponding resonator, $\psi(t)$, represents the eigenfunction basis plus the bias Δ. This bias can be removed by subtracting Δ from $\psi(t)$.

Subsequently, using Equations (80) and (82) we can determine the resonator model frequency process for each resonator model so that the resonator is equivalent to the eigenfunction,

$$(2\pi f_j(t))^2 = -\frac{1}{\tilde{\phi}_j(t)}\frac{d^2\tilde{\phi}_j(t)}{dt^2} \ . \tag{83}$$

Figure 13 compares the eigenfunction and corresponding resonator whose frequency profiles are calculated using Equation (83). These basis functions are the four most significant eigenfunctions for the non-stationary periodic covariance function in Equation (81) with $D = 10$, $\alpha = 0.8$ and $l = 20$. The top panes show the eigenfunction and corresponding resonator and the bottom panes show the resonator coefficient, $(2\pi f(t))^2$, as per Equation (83), required by the resonator model to equate the resonator with the eigenfunction. We note the presence of negative resonator coefficient values $(2\pi f(t))^2$. These correspond to complex valued frequencies which model basis decay in a manner similar to the basis decay term in the alternative resonator model as per Equation (20).

We next examine the properties of the alternative resonator model, as per Equation (20), when representing perfectly periodic, non-stationary Gaussian processes. The alternative resonator model uses time invariant coefficients, A and B,

$$\frac{d^2\psi_j(t)}{dt^2} + A_j\frac{d\psi_j(t)}{dt} + B_j\psi_j(t) = \omega_j(t) \ , \tag{84}$$

where ω_j is a white noise component. Modelling the non-stationary process via Equation (84) avoids the need to compute frequency processes using the interacting multiple model (IMM) in the original formalisation of the resonator model (Särkkä et al., 2012). For perfectly periodic covariance functions (with period D) then $\psi_j(t + D) = \psi_j(t)$ for all t and consequently, as ω_j is i.i.d., then $\omega_j(t) = 0$ for all t. Thus, the solution of the previous equation is,

$$\psi_j(t) = G_j \exp\left[(\pm i\sqrt{B_j - 0.25A_j^2} - 0.5A_j)t\right] \ .$$

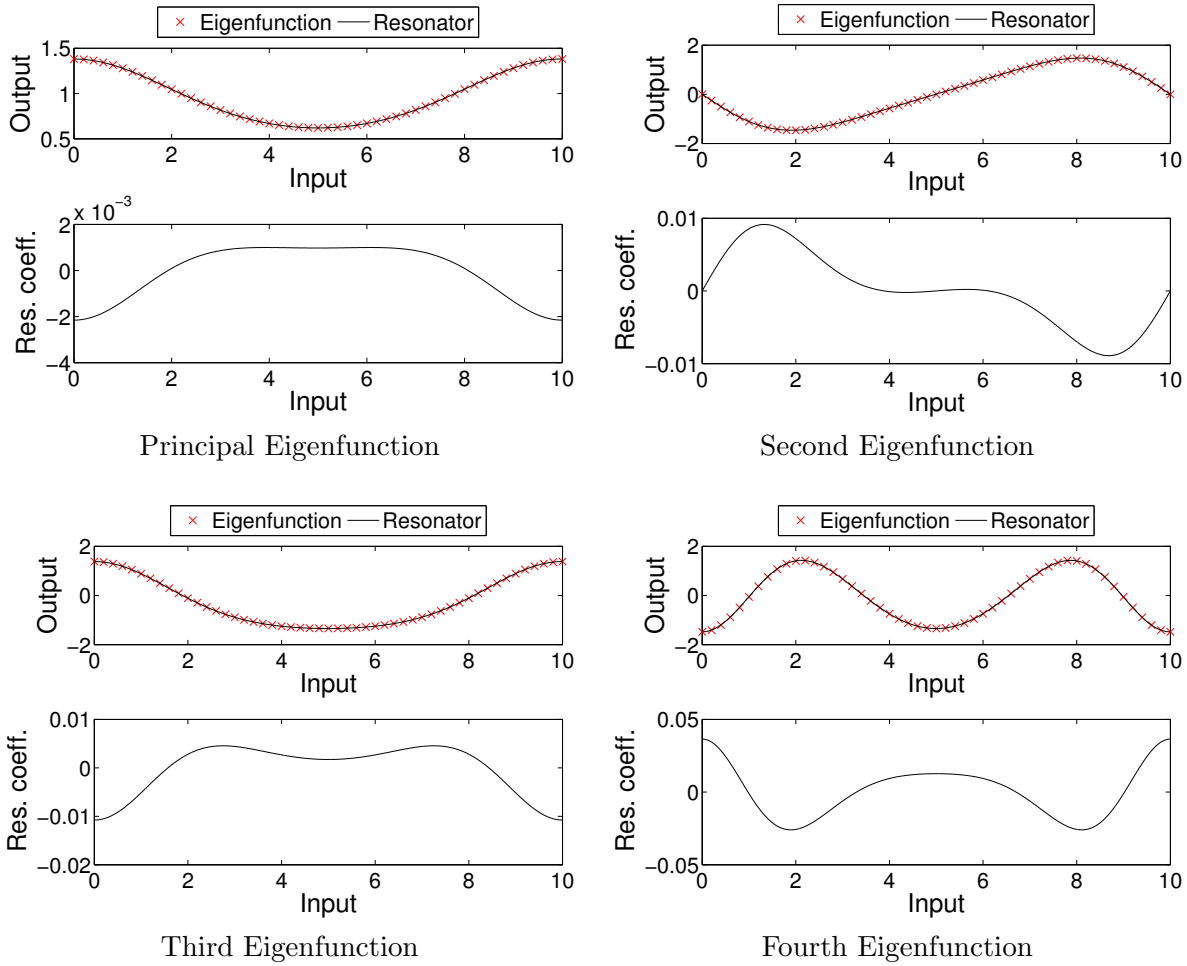
So that $\psi_j(t + D) = \psi_j(t)$ for all t then $A_j = 0$ and thus,

$$\psi_j(t) = G_j \exp\left[i\sqrt{B_j}t\right] \ .$$

Consequently, expanding the exponential in terms of cosine and sine functions we see that ψ_j must be the Fourier basis functions. The Fourier basis is a sub-optimal basis for non-stationary covariance functions as, in general, the optimal eigenfunction basis is not Fourier (see, for example, Figure 13). Thus, the resonator model, as per Equation (84), is a sub-optimal representation for non-stationary periodic covariance functions.

B.1.3 Quasi-Periodic Covariance Functions

A *quasi-periodic process* $g \sim \mathcal{GP}(b, K)$ is generated from a Gaussian process with covariance function of the form $K(t, t') = K_{\text{quasi}}(t, t')K_{\text{periodic}}(t, t')$ where K_{quasi} is non-periodic and



Principal Eigenfunction

Second Eigenfunction

Third Eigenfunction

Fourth Eigenfunction

Figure 13: The four most significant Eigenfunctions and coincident resonators for the non-stationary periodic covariance function in Equation (81). In each pane the top graph shows the eigenfunction and the lower graph shows the resonator coefficient $(2\pi f(t))^2$ profile required by the resonator model to equate the resonator with the eigenfunction.

K_{periodic} is perfectly periodic (either stationary or non-stationary). Equation (34) is an example of a quasi-periodic process covariance function. In general, when the periodic kernel, K_{periodic}, has period D then, with high probability, $g(t+D) \neq g(t)$ for all $t \in \mathbb{R}$ unlike the perfectly periodic case presented above.

Quasi-periodic eigenfunction models use a time varying weight coefficient, $a_j(t)$, as per Equation (35). Thus, extending Equation (79), in this case the resonator and eigenfunction linear basis models are equivalent if,

$$\psi_j(t) = a_j(t)\phi_j(t) \ .$$

Consequently, by substituting $\psi_j(t)$ into Equation (19), setting $\omega_j(t) = 0$ and rearranging we get the frequency process, $f_j(t)$, for each resonator for the quasi-periodic process,

$$(2\pi f_j(t))^2 = -\frac{\ddot{\phi}_j(t)a_j(t) + \phi_j(t)\ddot{a}_j(t) + 2\dot{\phi}_j(t)\dot{a}_j(t)}{a_j(t)\phi_j(t)} \ .$$

Since the coefficient process $a_j(t)$ is stochastic then so too is $f_j(t)$. We note that both $\psi_j(t)$ and $a_j(t)$ must be inferred when using the resonator model and the Kalman filter. This places significant computational cost on the Kalman filter prediction equations. Thus, we do not recommend implementing quasi-periodic GP priors with the resonator model as per Equation (19).

The alternative resonator model, as per Equation (20) encodes a decay term, via the first order derivative of the basis, appropriate for modelling quasi-periodic covariance functions. This model is investigated empirically in Section 9 on a home heating prediction problem which exploits quasi-periodic latent forces.

B.2 Computational Complexity of Eigenfunction and Resonator Models

When the Gaussian process covariance function for each latent force is known, so that we can generate the appropriate eigenfunction basis for any choice of covariance function hyperparameters, then searching over the hyperparameter values of the covariance function can be significantly less computationally demanding than searching over the frequency space for a potentially large number of resonators.

When constructing the eigenfunction model the greatest computational cost arises from calculating the Nyström approximation. However, the significant eigenfunctions can be found iteratively and efficiently using Von Mises iteration. At each iteration the next largest eigenvalue and corresponding eigenfunction are found. This approach continues until all the significant eigenvalues are found. If J eigenfunctions with the largest eigenvalues are found using Von Mises iteration then the complexity of our approach is $\mathcal{O}(JN^2)$ where $N \times N$ is the size of the Gram matrix in Equation (27) obtained by sampling the periodic covariance function. Inferring the eigenfunction model also involves searching over a relatively small set of p hyperparameters, often of the order of about $p = 3$ parameters comprising the input scale, output scale and the period of the covariance function. If the set of admissible values along each hyperparameter dimension has cardinality Υ then the computational complexity of searching the parameter space is $\mathcal{O}(\Upsilon^p)$. The overall computational complexity of inferring the eigenfunction model is therefore $\mathcal{O}(\Upsilon^p JN^2)$.

The parameters of the J-dimensional resonator model can be found by solving a non-convex optimisation problem over a $3J$ dimension parameter space where the parameters are J Fourier basis function frequencies, J basis function phases and J magnitudes for the basis power spectrum. If the set of admissible values along each dimension has cardinality Υ then the computational complexity of searching the parameter space is $\mathcal{O}(\Upsilon^{3J})$. To identify the optimal choice of parameter values each parameter vector constructed during the search over the parameter space requires the comparison of $\mathcal{O}(N^2)$ entries between the sampled kernel and the covariance matrix of the target function induced by the resonator model. Thus the resonator model is inferred with computational complexity $\mathcal{O}(\Upsilon^{3J}N^2)$. We note

that, whereas the resonator model training phase is exponentially complex in the number of basis functions, J, the eigenfunction model is linear in J. [11]

We compare the run times for the eigenfunction and resonator approaches empirically in Section 9.

B.3 Summary

We have demonstrated the link between the resonator model and the eigenfunction approach. Through this link we have been able to identify that,

1. the eigenfunction basis is optimal in that it minimises the mean squared error between the J-dimensional model and the target function.

2. the variant frequency term in the resonator second order differential equation provides sufficient flexibility to yield basis functions which are equivalent to the eigenfunctions.

3. we have developed an algorithm for deriving optimal resonator models for all perfectly periodic covariance functions from the eigenfunctions of the covariance function. Thus, we are able to offer an efficient mechanism for encoding the GP prior in the resonator model.

11. Note that, when using the Nyström approximation of the eigenfunctions it is necessary to store the eigenvectors from which the eigenfunctions can be calculated. Although, for the stationary case, they can be calculated when required from the cosine and sine functions.