POLYPUBLIE

A kernel-based method for modeling non-harmonic periodic phenomena in bayesian dynamic linear models

Luong Ha Nguyen, Ianis Gaudot, Shervin Khazaeli, James Alexandre Goulet

Document en libre accès dans PolyPublie
Open Access document in PolyPublie

Date	2019
Type	Article de revue / Article
Version	Version officielle de l'éditeur / Published version
URL de PolyPublie PolyPublie URL	https://publications.polymtl.ca/5056/
Évaluation par les pairs Peer reviewed	Révisé par les pairs / Refereed
Conditions d'utilisation Terms of use	Creative Commons Attribution 4.0 International (CC BY)
Référence Citation	Nguyen, L. H., Gaudot, I., Khazaeli, S., & Goulet, J. A. (2019). A kernel-based method for modeling non-harmonic periodic phenomena in bayesian dynamic linear models. Frontiers in Built Environment, 5, 8. https://doi.org/10.3389/fbuil.2019.00008

Publication officielle
Official publication

Titre de la revue Journal title	Frontiers in Built Environment (vol. 5)
Maison d'édition Publisher	Frontiers
URL officielle Official URL	https://doi.org/10.3389/fbuil.2019.00008
Mention légale Legal notice	Copyright © 2019, Nguyen, L. H., Gaudot, I., Khazaeli, S., & Goulet, J. A. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Ce fichier a été téléchargé à partir de PolyPublie, le dépôt institutionnel de Polytechnique Montréal
This file has been downloaded from PolyPublie, the institutional repository of Polytechnique Montréal
https://publications.polymtl.ca

frontiers in Built Environment

METHODS
published: 04 February 2019
doi: 10.3389/fbuil.2019.00008

A Kernel-Based Method for Modeling Non-harmonic Periodic Phenomena in Bayesian Dynamic Linear Models

Luong Ha Nguyen, Ianis Gaudot, Shervin Khazaeli and James-A. Goulet*

Department of Civil, Geologic and Mining Engineering, Polytechnique Montreal, Montreal, QC, Canada

Modeling periodic phenomena with accuracy is a key aspect to detect abnormal behavior in time series for the context of Structural Health Monitoring. Modeling complex non-harmonic periodic pattern currently requires sophisticated techniques and significant computational resources. To overcome these limitations, this paper proposes a novel approach that combines the existing Bayesian Dynamic Linear Models with a kernel-based method for handling periodic patterns in time series. The approach is applied to model the traffic load on the Tamar Bridge and the piezometric pressure under a dam. The results show that the proposed method succeeds in modeling the stationary and non-stationary periodic patterns for both case studies. Also, it is computationally efficient, versatile, self-adaptive to changing conditions, and capable of handling observations collected at irregular time intervals.

Keywords: Bayesian, dynamic linear models, kernel regression, structural health monitoring, kalman filter, dam, bridge

OPEN ACCESS

Edited by:
Ersin Aydin,
Niğde Ömer Halisdemir University, Turkey

Reviewed by:
Dimitrios Giagopoulos,
University of Western Macedonia, Greece
Yves Reuland,
École Polytechnique Fédérale de Lausanne, Switzerland

****Correspondence:***
Luong Ha Nguyen
luongha.nguyen@gmail.com

Specialty section:
This article was submitted to Structural Sensing, a section of the journal Frontiers in Built Environment

Received: *31 October 2018*
Accepted: *16 January 2019*
Published: *04 February 2019*

Citation:
Nguyen LH, Gaudot I, Khazaeli S and Goulet J-A (2019) A Kernel-Based Method for Modeling Non-harmonic Periodic Phenomena in Bayesian Dynamic Linear Models. Front. Built Environ. 5:8. doi: 10.3389/fbuil.2019.00008

1. INTRODUCTION

In civil engineering, *Structural Health Monitoring* (SHM) is a field of study that aims at monitoring the state of a civil structure based on time series data. One key aspect of SHM is to detect anomalies from the analysis of structural responses, including—but not limited to—displacement, velocity, acceleration, and strain measurements collected over time using, for example, laser displacement sensors, velocimeters, accelerometers, or strain gauges. Monitoring operates under real field conditions and, therefore, SHM must deal with observations collected at irregular time intervals, missing values, high noise level, sensor drift, and outliers (Goulet and Koo, 2018).

One applicable approach for SHM time series analysis and forecasting is to decompose the observed structural response into a set of generic components including baseline, periodic, and residual components (Léger and Leclerc, 2007; Goulet and Koo, 2018). The baseline component captures the baseline response of a civil structure, which is directly related to the irreversible changes due to the aging of a structure over time. The periodic components model the periodic reversible changes due to the influence of external effects such as temperature, humidity, pressure, and traffic load. The residual component takes into account the modeling error. Each component is associated with a set of unknown model parameters which needs to be learned from a training dataset.

Periodicity in time series data is very common. Periodicity is of major importance in SHM because the changes in magnitude of the periodic pattern due to external effects are usually much larger than the magnitude of irreversible changes related to the health of the structure. It is worth mentioning that modeling external effects, when they are observed, is of major importance.

Therefore, modeling periodic phenomena with accuracy is crucial to detect abnormal behavior in structural responses. The periodicity observed in structural responses have many origins: (i) annual cycles due to the seasons, (ii) daily cycles due to the day/night alternance, and (iii) hourly, weekly, and yearly cycles due to working days, weekends, and holidays (Goulet and Koo, 2018). In this paper, the terms periodicity (period), seasonality (season), and cyclicity (cycle) are used interchangeably whether or not the observed periodic patterns are related to well-known seasons. Periodic patterns found in SHM time series data may exhibit simple harmonic patterns, in the sense that they can be modeled using a single sinusoidal function. This is often the case for the effects induced by the daily and yearly air temperature variation. However, other periodic patterns are more complex. They are usually referred to as non-harmonic patterns because they cannot be modeled by a single sinusoidal function. For instance, in the field of dam engineering, the ambient air and water reservoir contribute to the establishment of a thermal gradient through the structure, which leads to a periodic non-harmonic pattern on the observed displacements time series (Léger and Leclerc, 2007; Tatin et al., 2015; Salazar and Toledo, 2018). In bridge monitoring, the effects related to traffic load variations generally display a non-harmonic weekly periodic pattern due to the day/night as well as the weekdays/weekend alternance (Westgate et al., 2015; Goulet and Koo, 2018).

Regression techniques, such as multiple linear regression (Léger and Leclerc, 2007; Tatin et al., 2015; Gamse and Oberguggenberger, 2017) and artificial neural networks (Mata, 2011), are widely used to perform time series decomposition in SHM. In the multiple linear regression approach, a sum of harmonic functions must be used to fit periodic non-harmonic patterns, whereas in the neural network approach, interconnected hidden layers are used to achieve such a purpose. Regression techniques have several limitations, as they are known to be sensitive to the outlier, to be prone to overfitting, and to require a large amount of available data to learn the model parameters (Murphy, 2012). In addition, regression approaches are not capable to self-adapt to changing conditions without retraining the model. This is a considerable downside, as SHM time series are usually non-stationary. Here, the term stationary can be generally referred to as not evolving through time. For instance, a change in the trend of the structure baseline response due to a damage, or a change in the amplitude of the periodic pattern due to an extreme climatic event, lead to non-stationarity.

Bayesian Dynamic Linear Model (West and Harrison, 1999), hereafter referred to as BDLM, is a class of state-space model (SSM) that allows non-stationary components to be learned without retraining the model. In the context of SSM, the components are known as hidden states because they are not directly observed. During the last decades, BDLMs have been used in many disciplines concerned with time series analysis and forecasting (West and Harrison, 1999; Särkkä, 2013), and recently, it has been shown that BDLM is capable of tracking non-stationary baseline of observed structural responses for bridge monitoring purposes (Solhjell, 2009; Goulet, 2017; Goulet and Koo, 2018). BDLMs have also been used to detect abnormal behavior in the response structure of a dam (Nguyen and Goulet, 2018a). The Fourier component representation (West and Harrison, 1999), the autoregressive component representation (Prado and West, 2010), and the dynamic regression approach (Nguyen and Goulet, 2018b) are applicable techniques to handle periodic components in BDLMs. Such approaches are capable of modeling harmonic and simple non-harmonic pattern, but they are not well-suited for handling complex non-harmonic periodic phenomena. A review of existing methods capable of modeling non-harmonic periodic phenomena is presented in section 2.4.

Kernel-based methods have gained much attention in recent years, mainly due to the high performance that they provide in a variety of tasks (Scholkopf and Smola, 2001; Álvarez et al., 2012). Kernel regression is a non-parametric approach that uses a known function (the kernel function) to fit non-linear patterns in the data (Henderson and Parmeter, 2015). The use of kernel enables to interpolate between a set of so-called *control points* as well as to extrapolate beyond them, which is well suited for handling non-uniformly spaced observations over time. The periodic kernel regression approach (Michalak, 2011) utilizes a periodic kernel function to model periodic patterns. Nevertheless, periodic kernel regression is a regression approach in which the model does not evolve over time, and therefore it cannot be used for processing SHM time series data.

This study proposes to combine the periodic kernel regression and the BDLM approaches to model complex non-harmonic periodic phenomena in SHM time series data. The new technique is illustrated on real time series data recorded for bridge and dam monitoring purposes. The paper is organized as follows. The first section reviews the BDLM theory, by focusing on the approaches already available in the literature to model periodic phenomena in the SSM framework. Then, after describing the main features of the kernel regression, the third section explains how periodic kernel regression can be coupled with the existing BDLM formulation to model non-harmonic periodic patterns. Section 5 presents two examples of applications from real case studies.

2. BAYESIAN DYNAMIC LINEAR MODELS

BDLMs consists in a particular type of SSM where the hidden states of a system are estimated recursively through time. The estimation combines the information coming from a linear dynamic model describing the transitions of the states across time as well as the information coming from empirical observations. The key aspect of BDLM is to provide a set of generic components which form the building blocks that can be assembled for modeling a wide variety of time series. The current section reviews the mathematical formulation for SSM, presents the specific formulation for BDLM periodic components, and explains how to learn the parameters involved in these formulations.

2.1. Mathematical Formulation

The core of BDLMs like other state-space models is the *hidden state variables* $\mathbf{x}_t = [x_1, x_2, \cdots, x_X]^\intercal$, where the term hidden refers to the fact that these variables are not directly observed.

The evolution over time of hidden state variables is described by the linear *transition model*

$$\mathbf{x}_t = \mathbf{A}_t\mathbf{x}_{t-1} + \mathbf{w}_t,\ \mathbf{w}_t \sim \mathcal{N}(\mathbf{w};\mathbf{0},\mathbf{Q}_t), \tag{1}$$

which is defined by the transition matrix $\mathbf{A}_t$ and process noise $\mathbf{Q}_t$. The *observation model*

$$\mathbf{y}_t = \mathbf{C}_t\mathbf{x}_t + \mathbf{v}_t,\ \mathbf{v}_t \sim \mathcal{N}(\mathbf{v};\mathbf{0},\mathbf{R}_t) \tag{2}$$

is defined by the observation matrix $\mathbf{C}_t$ and the observation covariance $\mathbf{R}_t$. The observation model defines which of the hidden variables in $\mathbf{x}_t$ contribute to the observations $\mathbf{y}_t$. In BDLM, the formulation of $\{\mathbf{A}_t, \mathbf{C}_t, \mathbf{Q}_t, \mathbf{R}_t\}$ matrices is defined for generic model components which can model behavior such as a locally constant behavior, a trend, an acceleration, a harmonic periodic evolution or an autoregressive behavior. For a complete review of these components, the reader should consult references by West and Harrison (1999), Murphy (2012), and Goulet (2017).

2.2. Hidden State Estimation

The estimation of hidden state variables $\mathbf{x}_t$ is intrinsically dynamic and sequential, which requires defining a special shorthand notation. First, $y_{1:t} \equiv \{y_1, y_2, \cdots, y_t\}$ defines a set of observations ranging from time 1 to t. $\boldsymbol{\mu}_{t|t} \equiv \mathbb{E}[\mathbf{X}_t|y_{1:t}]$ defines the expected value of a vector of hidden states at time t conditioned on all the data available up to time t. In the case where the data at the current time t is excluded, the notation changes for $\boldsymbol{\mu}_{t|t-1} \equiv \mathbb{E}[\mathbf{X}_t|y_{1:t-1}]$. Finally, $\boldsymbol{\Sigma}_{t|t} \equiv \text{var}[X_t|y_{1:t}]$ defines the covariance of a vector of hidden states at time t conditioned on all the data available up to time t. The recursive process of estimating hidden state variables $\mathbf{x}_t$ from the knowledge at $t-1$ is summarized by

$$(\boldsymbol{\mu}_{t|t}, \Sigma_{t|t}, \mathcal{L}_t) = \text{Filter}(\boldsymbol{\mu}_{t-1|t-1}, \Sigma_{t-1|t-1}, y_t, \mathbf{A}_t, \mathbf{C}_t, \mathbf{Q}_t, \mathbf{R}_t), \tag{3}$$

where $\mathcal{L}_t$ is the joint marginal log-likelihood presented in section 2.3 and the Filter can be performed using, for examples, the *Kalman* or the *UD* filter (Simon, 2006).

2.3. Parameter Estimation

Matrices $\{\mathbf{A}_t, \mathbf{C}_t, \mathbf{Q}_t, \mathbf{R}_t\}$ involve several parameters to be learned using either a Bayesian or a maximum likelihood estimation (MLE) technique. In this paper, we focus solely on the MLE approach where the objective is to identify the vector of optimal parameters

$$\boldsymbol{\theta}^* = \arg\max_{\boldsymbol{\theta}} \ln f(\mathbf{y}_{1:T}|\boldsymbol{\theta}), \tag{4}$$

which maximizes the joint log-likelihood. The joint likelihood is obtained as the product of the *marginal likelihood* at each time t

$$f(\mathbf{y}_{1:T}|\boldsymbol{\theta}) = \prod_{t=1}^{T} f(\mathbf{y}_t|\mathbf{y}_{1:t-1}, \boldsymbol{\theta}). \tag{5}$$

The marginal likelihood that corresponds to the joint prior probability of observations is defined as

$$f(\mathbf{y}_t|\mathbf{y}_{1:t-1}, \boldsymbol{\theta}) = \mathcal{N}\left(\mathbf{y}_t; \mathbf{C}_t(\boldsymbol{\theta})\boldsymbol{\mu}_{t|t-1}, \mathbf{C}_t(\boldsymbol{\theta})\Sigma_{t|t-1}\mathbf{C}_t(\boldsymbol{\theta})^\mathsf{T} + \mathbf{R}_t(\boldsymbol{\theta})\right). \tag{6}$$

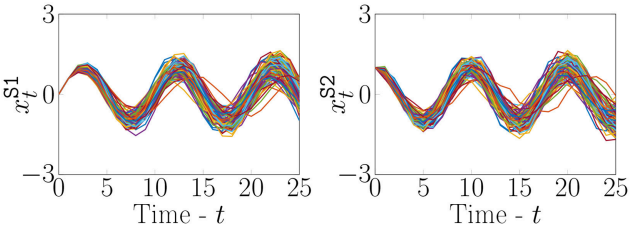


FIGURE 1 | Examples of realizations of x^{S} for a Fourier Form process with parameters $\sigma_w^S = 0.01$ and $p = 10$.

The maximization task in Equation (4) can be performed with a variety of gradient-based optimization algorithms such as *Newton-Raphson* or *Stochastic Gradient Ascent* (Bishop, 2006).

2.4. Modeling Periodic Phenomena

Several generic components for modeling periodic phenomena are already available for the existing BDLM. The following subsection reviews the strengths and limitations of the three existing generic components.

2.4.1. Fourier Form

The *Fourier Form* component allows to model sine-like phenomena. Its mathematical formulation follows

$$\mathbf{x}_t^{\mathrm{S}} = \begin{bmatrix} x_t^{\mathrm{S1}} \\ x_t^{\mathrm{S2}} \end{bmatrix},\ \mathbf{A}_t^{\mathrm{S}} = \begin{bmatrix} \cos\omega & \sin\omega \\ -\sin\omega & \cos\omega \end{bmatrix},\ \mathbf{C}_t^{\mathrm{S}} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}^\mathsf{T},$$
$$\mathbf{Q}_t^{\mathrm{S}} = (\sigma_w^{\mathrm{S}})^2 \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \tag{7}$$

where $\omega = \frac{2\pi\cdot\Delta t}{p}$ is the angular frequency defined by the period p of the phenomena modeled, and Δt, the time-step length. **Figure 1** presents examples of realizations of a Fourier process.

Although the Fourier form component is computationally efficient, it is limited to modeling simple harmonic periodic phenomena such as the one presented in **Figure 2A**. It is possible to model simple non-harmonic periodic phenomena such as the one presented in **Figure 2B** using a superposition of Fourier form components each having a different period. Nevertheless this process is in practice difficult, especially when the complexity of the periodic pattern increases because it requires a significant increase in the number of hidden state variables as well as of unknown periods to be identified, which in turn decreases the overall computationally efficiency.

2.4.2. Dynamic Regression

In order to overcome the limitations of the Fourier form component, Nguyen and Goulet (2018b) have proposed a new formulation based on the *Dynamic Regression* component. This approach employs a cubic spline as depicted in **Figure 3** for modeling non-harmonic periodic phenomena such as the one presented in **Figure 2B**. The formulation for this component is

$$\mathbf{x}_t^{\mathrm{DR}} = x_t^{\mathrm{DR}},\ \mathbf{A}_t^{\mathrm{DR}} = 1,\ \mathbf{Q}_t^{\mathrm{DR}} = (\sigma_w^{\mathrm{DR}})^2,\ \mathbf{C}_t^{\mathrm{DR}} = g(t, \boldsymbol{\theta}^{\mathrm{DR}}). \tag{8}$$

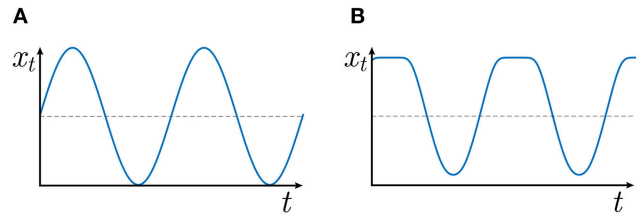


FIGURE 2 | Examples of harmonic **(A)** and non-harmonic **(B)** periodic pattern.

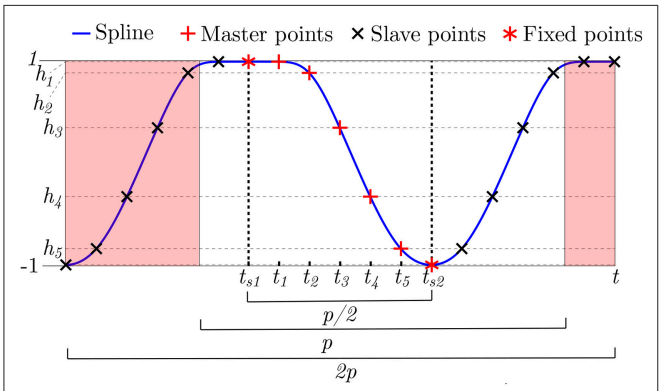


FIGURE 3 | Illustration of the spline employed with the Dynamic regression formulation proposed by Nguyen and Goulet (2018b).

With the dynamic regression component, the hidden state variable x_t^{DR} acts as a regression coefficient that is modulating the amplitude of $g(t, \boldsymbol{\theta}^{\mathrm{DR}})$ which corresponds to the spline presented in **Figure 3**. The parameters involving this spline are defined as $\boldsymbol{\theta}^{\mathrm{DR}} = [t_{s1}, p, \mathbf{h}]^\intercal$, $\mathbf{h} = [h_1, \cdots, h_{\mathrm{N}}]^\intercal$, where $\left\{\boldsymbol{\theta}^{\mathrm{DR}}, \mathbf{h}\right\}$ are learned from data. Although this method has been shown to be capable of modeling complex periodic patterns, its main limitation is that complex periodic patterns typically require a large number (N) of control-point parameters making the approach computationally demanding. The reason behind this computational demand is that control points $\mathbf{h}$ need to be estimated with an optimization algorithm (see section 2.3) rather than being treated as hidden state variables that are estimated using the filtering procedure introduced in section 2.2.

2.4.3. Non-parametric Periodic Component

Another alternative to model periodic phenomena is to employ the *non-parametric periodic component* (West and Harrison, 1999). The mathematical formulation for this component is

$$\mathbf{x}_t^{\mathrm{NP}} = [x_{t,1}^{\mathrm{NP}}, \cdots, x_{t,\mathrm{N}}^{\mathrm{NP}}]^\intercal,\ \mathbf{A}_t^{\mathrm{NP}} = \begin{bmatrix} \mathbf{0} & \mathbf{I}_{\mathrm{N}-1} \\ 1 & \mathbf{0} \end{bmatrix},\ \mathbf{Q}_t^{\mathrm{NP}} = \begin{bmatrix} (\sigma_w^{\mathrm{NP}})^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0}_{\mathrm{N}} \end{bmatrix},$$
$$\mathbf{C}_t^{\mathrm{NP}} = [1, 0, \cdots, 0], \tag{9}$$

where $\mathbf{x}_t^{\mathrm{NP}}$ is a vector of hidden state variables that are permuted at each time step by the transition matrix $\mathbf{A}_t^{\mathrm{NP}}$. The advantage of this approach in comparison with the dynamic regression technique is that it can model complex non-harmonic periodic phenomena while having no parameters to be estimated through optimization. This is possible because the signal is modeled using hidden state variables which are estimated recursively. The main limitation of this approach is that it requires as many hidden state variables as there are time steps within the period p of the phenomenon studied. For example, if the phenomenon studied have a period of 1 year and the data's acquisition frequency is $\Delta t = 30\,\mathrm{min}$, it results in $365 \times 48 = 17,520$ hidden state variables which is computationally inefficient.

3. KERNEL REGRESSION

The new method that employs the concepts of Kernel regression (Henderson and Parmeter, 2015) allows modeling periodic phenomena while overcoming the limitations of the methods presented in section 2.4. This section introduces the theory of Kernel regression on which this paper builds in the following section.

The *kernel* is employed to measure the similarity between pairs of covariates. The goal here is to model the periodic phenomena in time series, therefore the periodic kernel (Duvenaud, 2014) is formulated as

$$k(t_i, t_j) = \exp\left[-\frac{2}{\ell^2}\sin\left(\pi\frac{t_i - t_j}{p}\right)^2\right], \tag{10}$$

where the covariate is the time t. The kernel output $k(t_i, t_j) \in (0, 1)$ measures the similarity between two timestamps t_i and t_j as a function of the distance between these, as well as a function of two parameters; the period and kernel lengthscale, $\boldsymbol{\theta} = \{p, \ell\}$. **Figure 4** presents three examples of periodic kernels for different sets of parameters $\boldsymbol{\theta}$. With kernel regression we consider we have a set of observations $\mathcal{D} = \{(t_i, y_i), \forall i = 1 : \mathrm{D}\}$ consisting in D pairs of observed system responses y_i, each associated with its time of occurrence t_i. The regression model is built using a set of control points defined using $\mathbf{t}^{\mathrm{KR}} \in \mathbb{R}^{\mathrm{N}}$, a vector of time covariates associated with a vector of hidden control point values $\mathbf{x} \in \mathbb{R}^{\mathrm{N}}$. The observed system responses are modeled following

$$y_i = \mathbf{x}^\intercal \frac{k(t_i, \mathbf{t}^{\mathrm{KR}})}{\sum_t k(t, \mathbf{t}^{\mathrm{KR}})} + v_i,\quad \underbrace{v : V \sim \mathcal{N}(v; 0, \sigma_v^2)}_{\text{Observation errors}}, \tag{11}$$

where the predictive capacity of the model comes from the product of the hidden control point values $\mathbf{x}$ and the normalized kernel values which measure the similarity between the time t_i and those stored for the control points in $\mathbf{t}^{\mathrm{KR}}$. The main challenge here is to estimate $\mathbf{x}$ using the set of observation $\mathcal{D}$. The likelihood describing the conditional probability of $\mathcal{D}$ given $\mathbf{x}$ is

$$f(\mathcal{D}|\mathbf{x}) = \prod_{i=1}^{\mathrm{D}} \mathcal{N}\left(y_i; \mathbf{x}^\intercal \frac{k(t_i, \mathbf{t}^{\mathrm{KR}})}{\sum_t k(t, \mathbf{t}^{\mathrm{KR}})}, \sigma_v^2\right). \tag{12}$$

If like for the case presented in section 2.3 one employs a MLE approach to estimate $\mathbf{x}^*$, the problem consists in maximizing the log-likelihood following

$$\mathbf{x}^* = \arg\max_{\mathbf{x}} \ln f(\mathcal{D}|\mathbf{x}). \tag{13}$$

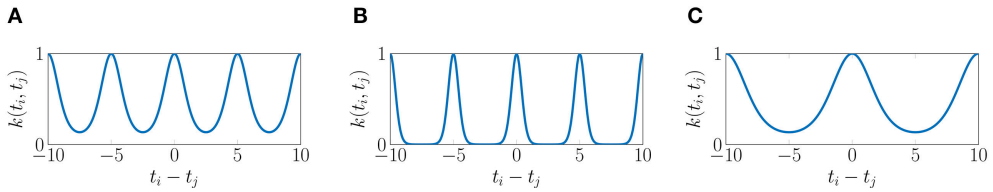


FIGURE 4 | Examples of periodic kernels. **(A)** $\boldsymbol{\theta} = \{p = 5, \ell = 1\}$, **(B)** $\boldsymbol{\theta} = \{p = 5, \ell = 0.5\}$, **(C)** $\boldsymbol{\theta} = \{p = 10, \ell = 1\}$.

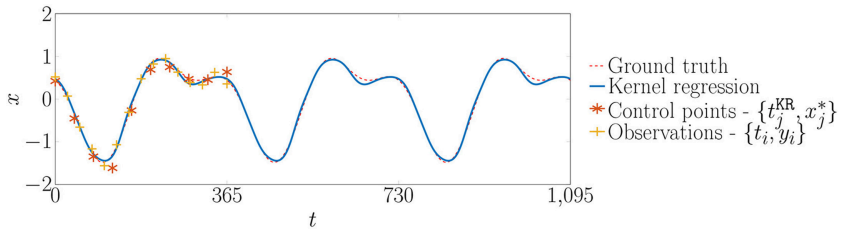


FIGURE 5 | Examples of application of kernel regression.

Figure 5 presents an example of application of the Kernel regression in order to model a non-harmonic periodic phenomena. A set of $\mathtt{D} = 15$ simulated observations represented by $+$ symbols are generated by adding normal-distributed observation noise on the ground truth signal presented by the dashed line. Then, an optimization algorithm is employed in order to identify the optimal values $\mathbf{x}^*$ for a set $\mathtt{N} = 10$ control points represented by the $*$ symbols. This example illustrates the capacity of the kernel-based method on interpolating the system response within control points as well as on extrapolating beyond them.

The main limitation of this approach is that the model cannot involve over time. Moreover, like the Dynamic Regression method presented in section 2.4.2, the estimation of $\mathbf{x}^*$ relies on a maximization algorithm which makes it computationally inefficient when applied to complex patterns.

4. USING KERNELS WITH BDLM

This section presents the new method proposed in this paper which consists in coupling the Kernel regression method reviewed in section 3 with the Dynamic regression method reviewed in section 2.4.2.

For the new approach, we consider that a vector of $\mathtt{N}+1$ hidden control point values $\mathbf{x}_t^{\mathrm{KR}} = [x_{t,0}^{\mathrm{KR}}, x_{t,1}^{\mathrm{KR}}, \cdots, x_{t,\mathrm{N}}^{\mathrm{KR}}]^\intercal$ is the vector of *hidden state variables* to be estimated using the filtering procedure presented in section 2.2. Therefore, the transition matrix is defined as

$$\mathbf{A}_t^{\mathrm{KR}} = \begin{bmatrix} 0 & \tilde{\boldsymbol{k}}^{\mathrm{KR}}(t, \mathbf{t}^{\mathrm{KR}}) \\ \mathbf{0} & \mathbf{I}_{\mathrm{N}} \end{bmatrix}, \tag{14}$$

where $\tilde{\boldsymbol{k}}^{\mathrm{KR}}(t, \mathbf{t}^{\mathrm{KR}})$ corresponds to the normalized kernel, $k(t, \mathbf{t}^{\mathrm{KR}}) / \sum_t k(t, \mathbf{t}^{\mathrm{KR}})$, as presented in Equation (10). $\tilde{\boldsymbol{k}}^{\mathrm{KR}}(t, \mathbf{t}^{\mathrm{KR}})$ is parameterized by the kernel lengthscale ℓ^{KR}, its period p^{KR}, and a vector of $\mathtt{N}$ timestamps $\mathbf{t}^{\mathrm{KR}} = [t_1^{\mathrm{KR}}, \cdots, t_{\mathrm{N}}^{\mathrm{KR}}]$, so that

$$\tilde{\boldsymbol{k}}^{\mathrm{KR}}(t, \mathbf{t}^{\mathrm{KR}}) \equiv \tilde{\boldsymbol{k}}^{\mathrm{KR}}(t, \mathbf{t}^{\mathrm{KR}}, \ell^{\mathrm{KR}}, p^{\mathrm{KR}}) \equiv \tilde{\boldsymbol{k}}_t^{\mathrm{KR}}, \tag{15}$$

where each timestamp t_i^{KR} is associated with a hidden control point value $x_{t,i}^{\mathrm{KR}}$. The first hidden state variable $x_{t,0}^{\mathrm{KR}}$ describes the current predicted pattern value which is obtained by multiplying the remaining $\mathtt{N}$ state variables $[x_{t,1}^{\mathrm{KR}}, \cdots, x_{t,\mathrm{N}}^{\mathrm{KR}}]^\intercal$ and the normalized kernel values contained in the first line of $\mathbf{A}_t^{\mathrm{KR}}$. From now on, the first and remaining hidden state variables are called the *hidden predicted pattern* and *hidden control points*, respectively. The $\mathtt{N} \times \mathtt{N}$ identity matrix forming the bottom right corner of $\mathbf{A}_t^{\mathrm{KR}}$ indicates that each of the hidden state variables $[x_{t,1}^{\mathrm{KR}}, \cdots, x_{t,\mathrm{N}}^{\mathrm{KR}}]^\intercal$ evolves over time following a random walk model (West and Harrison, 1999). The temporal evolution of these hidden state variables is controlled by the process noise covariance matrix

$$\mathbf{Q}_t^{\mathrm{KR}} = \begin{bmatrix} (\sigma_0^{\mathrm{KR}})^2 & \mathbf{0} \\ \mathbf{0} & (\sigma_1^{\mathrm{KR}})^2 \cdot \mathbf{I}_{\mathrm{N}} \end{bmatrix}, \tag{16}$$

where σ_1^{KR} controls the increase in the variance of the hidden control points between successive time steps and σ_0^{KR} controls the time-independent process noise in the hidden predicted pattern. This process noise allows describing random, unpredictable changes in the periodic phenomena between successive time steps. There are four possible cases: (1: $\sigma_0^{\mathrm{KR}} = 0$, $\sigma_1^{\mathrm{KR}} = 0$) the hidden predicted pattern is stationary and can be exactly modeled by the kernel regression, (2: $\sigma_0^{\mathrm{KR}} > 0$, $\sigma_1^{\mathrm{KR}} = 0$) the hidden predicted pattern is stationary yet, the kernel regression is an approximation of the true process being modeled (3: $\sigma_0^{\mathrm{KR}} = 0$, $\sigma_1^{\mathrm{KR}} > 0$) the hidden predicted pattern is non-stationary so the hidden control points values evolve over time and it can be exactly modeled by the kernel regression, (4: $\sigma_0^{\mathrm{KR}} >$

0, $\sigma_1^{\text{KR}} > 0$) the hidden predicted pattern is non-stationary and it cannot be exactly modeled by the kernel regression. As mentioned, above only the predicted pattern contributes directly to the observation so the observation matrix is

$$\mathbf{C}_t^{\text{KR}} = [1, 0, \cdots, 0]. \tag{17}$$

Because all control points are considered as hidden state variables, the only parameters to be estimated using the MLE procedure presented in section 2.3 are $\boldsymbol{\theta}^{\text{KR}} = [\sigma_0^{\text{KR}}, \sigma_1^{\text{KR}}, \ell^{\text{KR}}, p^{\text{KR}}]^\intercal$.

5. EXAMPLES OF APPLICATION

This section presents two examples of applications using a stationary and a non-stationary case. The first one involves traffic data recorded on a bridge and the second one involves the piezometric pressure recorded under a dam in service.

5.1. Stationary Case—Traffic Load

5.1.1. Data Description

This case-study is conducted on the traffic loading data collected on the Tamar Bridge from September 01 to October 21, 2007 with a total of 2,409 data points (Goulet and Koo, 2018). **Figure 6** presents the entire dataset. The traffic loading data are recorded from a toll booth where the vehicles are counted and classified by weight classes. Note that the data are collected with a uniform time steps of 30 min and is measured by kiloton (kT) unit. A constant baseline and a periodic pattern with a period of 7 days can be observed from the raw data. The traffic loading on weekends is much lighter than those on weekdays. For most of the day, the traffic load presents a high volume between 8 a.m. and 4 p.m., yet it drops out after 20 p.m. at night. In order to examine the predictive performance, the data are divided into a *training set* (1,649 data points) and a *test set* (760 data points). The unknown model parameters are learned using the training set. The predictive performance is then evaluated using the test set. The test set is presented by the shaded region in **Figure 6**.

5.1.2. Model Construction

The model to interpret the traffic load data consists in a vector of hidden states that includes a baseline (B) component to the average of the traffic load, a kernel regression (KR) component with 101 hidden state variables to model the periodic pattern, and an autoregressive (AR) component to capture the time-dependent model errors. The vector of hidden states is defined as

$$\mathbf{x}_t = \left[x_t^{\text{B}}, x_{t,0}^{\text{KR}}, x_{t,1}^{\text{KR}}, \cdots, x_{t,100}^{\text{KR}}, x_t^{\text{AR}}\right]^\intercal. \tag{18}$$

The model involves a vector of unknown model parameters $\boldsymbol{\theta}$ that are defined following

$$\boldsymbol{\theta} = \left[\sigma_w^{\text{B}}, \sigma_{w,0}^{\text{KR}}, \sigma_{w,1}^{\text{KR}}, \ell^{\text{KR}}, p^{\text{KR}}, \phi^{\text{AR}}, \sigma_w^{\text{AR}}, \sigma_v\right]^\intercal, \tag{19}$$

where σ_w^{B} is the baseline standard deviation, $\sigma_{w,0}^{\text{KR}}$ is the standard deviation of the hidden predicted pattern, $\sigma_{w,1}^{\text{KR}}$ is the standard deviation for the hidden control points, ℓ^{KR} is the kernel lengthscale, p^{KR} is the kernel period, ϕ^{AR} is the autoregression coefficient, σ^{AR} is the autoregression standard deviation, and σ_v is the observation noise. The standard deviation, kernel lengthscale, and kernel period are positive real number, $\mathbb{R}^+$. Because the autoregressive component is assumed to be stationary (Rao, 2008), the autoregression coefficient is constrained to the interval $(0, 1)$. For an efficient optimization, the model parameters are transformed in the unbounded space (Gelman et al., 2014). A natural logarithm function is applied to the standard deviation, kernel lengthscale, and kernel period. A sigmoid function is employed for the autoregression coefficient. The complete model matrices are detailed in **Appendix A**. The initial parameter values in the original space are selected using expert judgment and experience as well as prior data analysis

$$\boldsymbol{\theta}^0 = \left[10^{-6}, 0.29, 0.029, 0.5, 7, 0.75, 0.075, 1.47\right]^\intercal. \tag{20}$$

5.1.3. Parameter and Hidden State Estimation

The optimal model parameters obtained using the MLE technique presented in section 2.3 are

$$\boldsymbol{\theta}^* = \left[10^{-6}, 3.3 \times 10^{-5}, 10^{-5}, 0.0511, 7, 0.78, 0.34, 1.2 \times 10^{-5}\right]^\intercal, \tag{21}$$

where the ordering of each model parameter remains identical as in Equation 19. Because $\sigma_{w,0}^{\text{KR}}$ and $\sigma_{w,1}^{\text{KR}}$, are near to zero, the periodic pattern for this case-study is stationary. **Figure 7** presents the hidden state variables and predicted means for the traffic load estimated using Kalman Smoother (Murphy, 2012) for both training set and test set. $\boldsymbol{\mu}_{t|T}$ and $\boldsymbol{\sigma}_{t|T}$ are the mean value and standard deviation at time t, respectively, for the hidden state variables. The mean value $\boldsymbol{\mu}_{t|T}$ and its uncertainty bounds $\boldsymbol{\mu}_{t|T} \pm \boldsymbol{\sigma}_{t|T}$, are presented by the solid line and shaded

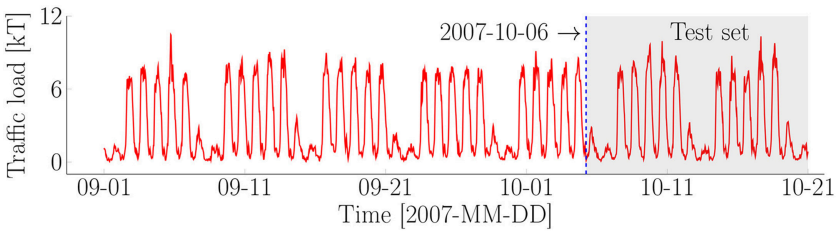


FIGURE 6 | Traffic load on the Tamar Bridge in the United Kingdom.

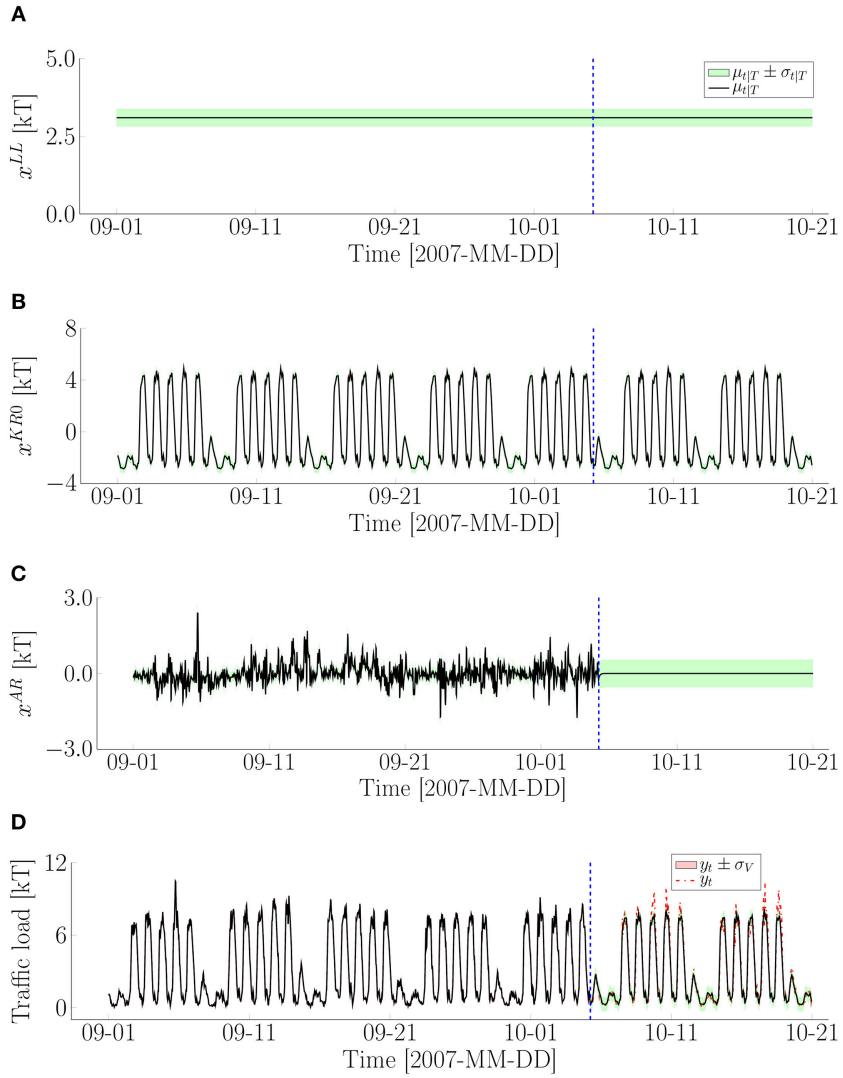


FIGURE 7 | Illustration of the estimation of hidden state variables for the traffic load data: **(A)** Baseline component, x_t^{B}; **(B)** Hidden predicted pattern, $x_{t,0}^{\mathrm{KR}}$; **(C)** Autoregressive component, x_t^{AR}; **(D)** Traffic load.

region. The estimates for the training set and test set are delimited by the dashed line. The traffic load data is presented by the dash-dot line. The results show that the autoregressive component, x_t^{AR}, presented in **Figure 7C**, is stationary as expected. Because the hidden predicted pattern is stationary, all cycles of 7 days have the identical pattern, as illustrated in **Figure 7B**. The estimates of the traffic load in the test set are visually close to the corresponding observations, as illustrated in **Figure 7D**. It can be seen that the model is not capable of predicting the peaks on the test set where the traffic loading presents a high volume in the rush hour. This can be explained by the high variability in the rush hour. The proposed method is intended to capture the periodic phenomena and not the changes from day to day. The *Mean Absolute Error* (MAE) and *Root Mean Squared Error* (RMSE) (Shcherbakov et al., 2013) are employed to measure the forecast accuracy by comparing the estimates with their corresponding traffic-load data in the test set. The mathematical formulation for the MAE and RMSE is presented in **Appendix C**. These metrics are evaluated on the forecast periods of 1, 3, 7, and 14 days in the test set. The results are summarized in **Table 1**. The small forecast errors are found for both metrics in each forecast period where the traffic loadings range from 0.07 to 10.3 kT. The forecast error increases with respect to the forecast period.

5.2. Non-stationary Case—Piezometer Under a Dam

5.2.1. Data Description

For this case-study, the new approach is applied to the piezometric pressure recorded from 2007 to 2016 under a dam in Canada. Note that the piezometric pressure is measured by *water meters* (wm) unit. The entire dataset with a total of 4181 data points, is presented in **Figure 8**. The data is recorded with a

non-uniform time-step length, as illustrated in **Figure 9**. The time-step length varies in the range from 1 to 456 h, in which the most frequent time-steps is 12 h. A descending trend and a periodic pattern can be observed from the raw data. Note that the data points for the training set and test set are 2, 538 and 1, 643, respectively.

5.2.2. Model Construction

The vector of hidden states being used for the model in this case-study includes a baseline (B) component, a local trend (T) component, a kernel regression components (KR), and an autoregressive component (AR). The baseline and local trend components are used for modeling the behavior of piezometric pressure. The kernel regression component associated with 21 hidden state variables, is employed to describe the periodic pattern. The time-dependent errors are modeled by the autoregressive component. The vector of hidden states are written as

$$\mathbf{x}_t = \left[x_t^{\mathrm{B}}, x_t^{\mathrm{T}}, x_{t,0}^{\mathrm{KR}}, x_{t,1}^{\mathrm{KR}}, \cdots, x_{t,20}^{\mathrm{KR}}, x_t^{\mathrm{AR}}\right]^{\intercal}. \tag{22}$$

TABLE 1 | Evaluation of forecast accuracy with respect to the different forecast periods for the traffic load data.

Metric	Forecast period			
	1 (day) 48 pts	3 (days) 144 pts	7 (days) 336 pts	14 (days) 760 pts
MAE	0.148	0.197	0.365	0.383
RMSE	0.209	0.278	0.551	0.582

The unknown model parameters relating to these hidden states are defined as

$$\boldsymbol{\theta} = \left[\sigma_w^{\mathrm{T}}, \sigma_{w,0}^{\mathrm{KR}}, \sigma_{w,1}^{\mathrm{KR}}, \ell^{\mathrm{KR}}, p^{\mathrm{KR}}, \phi^{\mathrm{AR}}, \sigma_w^{\mathrm{AR}}, \sigma_v\right]^{\intercal}, \tag{23}$$

where σ_w^{T} is the local-trend standard deviation and the remaining model parameters have the same with the bounds presented in Equation (19). These model parameters are also transformed in the unconstrained space for the optimization task. The full model matrices are presented in **Appendix B**. The initial parameter values in the original space are regrouped following

$$\boldsymbol{\theta}^0 = \left[10^{-6}, 0.37, 0.018, 0.5, 365.24, 0.75, 0.075, 0.018\right]^{\intercal}. \tag{24}$$

5.2.3. Parameter and Hidden State Estimation

The vector of optimal model parameters are obtained following

$$\boldsymbol{\theta}^* = \left[3.6 \times 10^{-6}, 0.047, 0.015, 0.485, 363.07, 0.91, 0.035, 4.5 \times 10^{-5}\right]^{\intercal}, \tag{25}$$

where the ordering of each model parameters is the same as in Equation 23. Because $\sigma_{w,0}^{\mathrm{KR}}$ and $\sigma_{w,1}^{\mathrm{KR}}$ are different from zero, the hidden predicted pattern is non-stationary. The hidden state variables and estimates for the piezometric pressure are shown in **Figure 10**. The mean values $\boldsymbol{\mu}_{t|T}$ and its uncertainty bounds $\boldsymbol{\mu}_{t|T} \pm \boldsymbol{\sigma}_{t|T}$ are presented by the solid line and shaded region, respectively. The dashed line is used to separate the estimates between the training set and test set. The dash-dot line presents the piezometric pressure data. According to the results, the autoregressive component, x_t^{AR}, presents a stationary behavior in the training set, as shown in **Figure 10C**. The non-stationarity in the hidden predicted pattern can be observed in **Figure 10B** where each cycle of 363 days is different from each other.

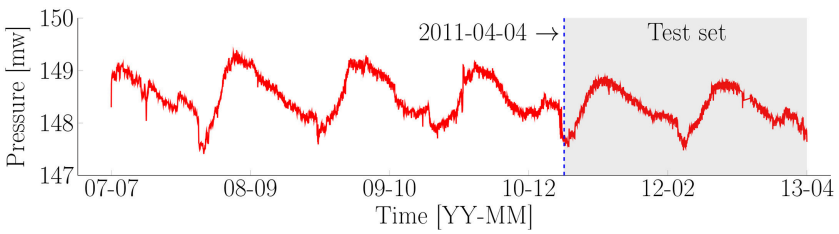


FIGURE 8 | Piezometric pressure under a dam in Canada.

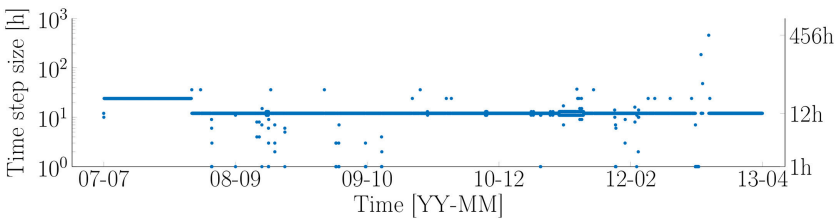


FIGURE 9 | Time-step length for the piezometric pressure data.

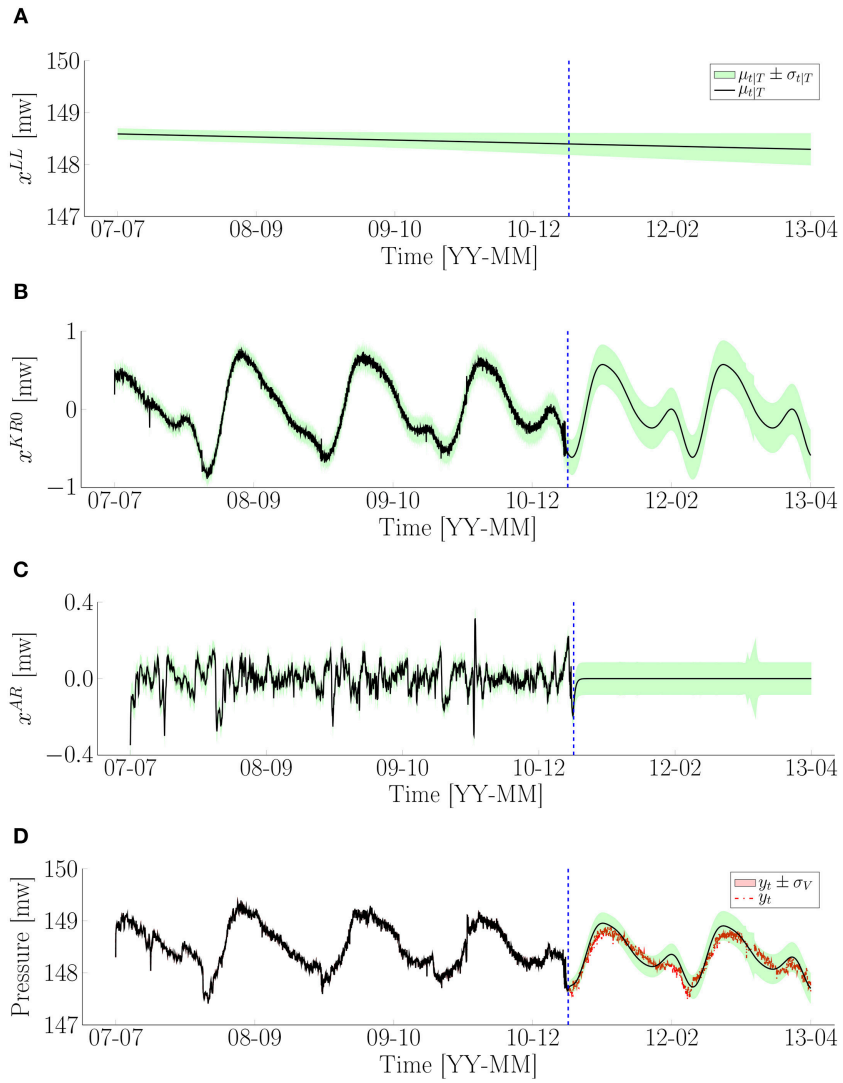


FIGURE 10 | Illustration of the estimation of hidden state variables for the piezometric pressure data: **(A)** Baseline component, x_t^{B}; **(B)** Hidden predicted pattern, $x_{t,0}^{\mathrm{KR}}$; **(C)** Autoregressive component, x_t^{AR}; **(D)** Pressure.

Figure 10D shows that the distance between the estimates for the first year in the test set and their corresponding observation is further than the one for the second year. Yet, the uncertainty bounds for the first year remain smaller than those for the second year. In addition, the uncertainty bounds of the estimates in the test set include most of the corresponding pressure data. The MAE and RMSE are also employed for measuring the forecast accuracy for such as this case-study. **Table 2** shows the forecast error of two metrics with respect to the forecast periods of 30, 180, 365, and 730 days. The forecast errors are small for each forecast period in which the piezometric pressure data varies in range from 147.5 to 148.9 water meters. The forecast error is large as the forecast period grows from 30 to 365 days. However, it is slightly reduced for the forecast period of 730 days. This can be explained by the better estimates for the second year in the test set.

6. DISCUSSION

The analysis of the traffic load and pressure time-series data shows that coupling the periodic kernel regression with BDLM enables to model non-harmonic periodic patterns recorded in different settings. The results from the traffic-load data allow investigating the performance of the approach to model a stationary periodic pattern observed regularly over time, but that exhibits a complex shape. The results from the pressure data allow interpreting the performance of the approach on a simple non-stationary periodic pattern observed irregularly over time.

In contrast with the dynamic regression technique presented in section 2.4.2, the proposed approach takes advantage of the Kalman computations to estimate analytically and recursively the periodic pattern. Therefore, the approach is computationally efficient. The use of Kalman computations also naturally provide

TABLE 2 | Evaluation of forecast accuracy for the different forecast periods on the piezometric pressure data.

Metric	Forecast period			
	30 (days) 62 pts	**180 (days) 357 pts**	**365 (days) 750 pts**	**730 (days) 1643 pts**
MAE	0.076	0.113	0.124	0.110
RMSE	0.0925	0.136	0.148	0.136

uncertainty bound around the prediction, which is crucial for time series forecasting. The use of a kernel in the proposed approach enables to process observations collected at irregular time interval, which is not possible with the non-parametric periodic component, as presented in section 2.4.3. In addition, it can handle non-stationary periodic patterns. The non-stationarity is accommodated by imposing a random walk on each hidden control points, which is mainly controlled by the standard deviation $\sigma_{w,1}^{\mathrm{KR}}$. In common cases, the time series cannot be visually inspected before analysis, and therefore, the model parameter $\sigma_{w,1}^{\mathrm{KR}}$ is learned from the data using the MLE procedure. Starting from $\sigma_{w,1}^{\mathrm{KR}}$ values of 0.029 and 0.18 for the traffic load and pressure data, respectively, the MLE procedure is capable to converge to $\sigma_{w,1}^{\mathrm{KR}}$ values of 10^{-5} and 0.015, respectively, thus highlighting the non-stationarity of the periodic pattern in the pressure data compared to the one observed in the traffic load data. The process noise standard deviation $\sigma_{w,0}^{\mathrm{KR}}$ associated with the hidden periodic pattern is also learned from the data. Initializing at $\sigma_{w,0}^{\mathrm{KR}}$ values of 0.29 and 0.37 for the traffic load and pressure data, respectively, the MLE procedure is capable to converge to $\sigma_{w,0}^{\mathrm{KR}}$ values of 3.3×10^{-5} and 0.047, respectively. The high values for $\sigma_{w,0}^{\mathrm{KR}}$ and $\sigma_{w,1}^{\mathrm{KR}}$ obtained from the pressure training dataset explain the shift in both amplitude and phase found in the estimates of the pressure test dataset, as illustrated in **Figure 10D**.

The kernel period, p^{KR}, is another crucial model parameter. Unlike the model parameters $\sigma_{w,0}^{\mathrm{KR}}$ and $\sigma_{w,1}^{\mathrm{KR}}$ discussed above, a good initial guess of the period should be provided for ensuring an efficient optimization. The reason is that the likelihood of the observation given the period is strongly peaked, and starting at the wrong value can lead to a slow convergence. Moreover, the likelihood about the period may also have several maxima if the time series contains several periods, which may lead the MLE procedure to converge toward a local maximum. The use of a periodogram for identifying the dominant period may be misleading, because the dominant period may not be the period that is the most suited for the kernel period. For example, the dominant period in the traffic loading example is 1 day, yet it can be observed that there is a difference in amplitude between the weekdays and weekends. Therefore, a period of 1 day might yield poor predictive results for such as case. For practical applications, the visualization of the time-series is the best way for identifying good starting values for the period prior analysis. In addition, multiple sets of initial parameter values should be tested during the optimization procedure.

Furthermore, the number of hidden control points is a hyperparameter to be tuned before the analysis. This hyperparameter can be identified using the cross-validation technique (Bergmeir and Benítez, 2012). The cross-validation consists in splitting the data into the training set and validation set in which the model is trained on the training set. The hyperparameter that minimizes the error on the validation set is selected. Moreover, it is important to note that the kernel function allows interpolating between hidden control points, thus enabling simple models (i.e., with few control points) to fit a large diversity of periodic patterns.

7. CONCLUSION

This study proposes a novel approach for modeling the non-harmonic periodic patterns in time-series for the context of Structural Health Monitoring. The technique combines the formulation of the existing Bayesian dynamic linear models with a kernel-based method. The application of the method on two case studies illustrates its potential for interpreting complex periodic time series. The results show that the proposed approach succeeds in modeling the stationary and non-stationary periodic patterns for both case studies. Also, it is computationally efficient, versatile, self-adaptive to changing conditions, and capable of handling observations collected at irregular time intervals.

AUTHOR CONTRIBUTIONS

LHN, IG, SK, and J-AG wrote the paper together and conceived the case-study examples. J-AG and LHN formulated the mathematical formulation for the proposed method. LHN and SK performed the simulation.

FUNDING

This project is funded by the Natural Sciences and Engineering Research Council of Canada (NSERC), Hydro Québec (HQ), Hydro Québec's Research Institute (IREQ), and Institute For Data Valorization (IVADO).

REFERENCES

Álvarez, M. A., Rosasco, L., and Lawrence, N. D. (2012). Kernels for vector-valued functions: a review. *Found. Trends Mach. Learn.* 4, 195–266. doi: 10.1561/2200000036

Bergmeir, C., and Benítez, J. M. (2012). On the use of cross-validation for time series predictor evaluation. *Information Sci.* 191, 192–213. doi: 10.1016/j.ins.2011.12.028

Bishop, C. (2006). *Pattern Recognition and Machine Learning*. Singapore: Springer.

Duvenaud, D. (2014). *Automatic Model Construction With Gaussian Processes*. Ph.D. thesis, University of Cambridge.

Gamse, S. and Oberguggenberger, M. (2017). Assessment of long-term coordinate time series using hydrostatic-season-time model for rock-fill embankment dam. *Struct. Control Health Monitor.* 24:e1859. doi: 10.1002/stc.1859

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2014). *Bayesian Data Analysis*, 3 Edn. New York, NY: CRC Press.

Goulet, J.-A. (2017). Bayesian dynamic linear models for structural health monitoring. *Struct. Control Health Monitor.* 24, 1545–2263. doi: 10.1002/stc.2035

Goulet, J.-A., and Koo, K. (2018). Empirical validation of Bayesian dynamic linear models in the context of structural health monitoring. *J. Bridge Eng.* 23:05017017. doi: 10.1061/(ASCE)BE.1943-5592.0001190

Henderson, D. J., and Parmeter, C. F. (2015). *Applied Nonparametric Econometrics*. Cambridge University Press.

Léger, P., and Leclerc, M. (2007). Hydrostatic, temperature, time-displacement model for concrete dams. *J. Eng. Mech.* 133, 267–277. doi: 10.1061/(ASCE)0733-9399(2007)133:3(267)

Mata, J. (2011). Interpretation of concrete dam behaviour with artificial neural network and multiple linear regression models. *Eng. Struct.* 33, 903–910. doi: 10.1016/j.engstruct.2010.12.011

Michalak, M. (2011). "Time series prediction with periodic kernels," in *Computer Recognition Systems 4*, eds R. Burduk, M. Kurzyński, M. Woźniak, and A. Żołnierek (Berlin; Heidelberg: Springer Berlin Heidelberg), 137–146.

Murphy, K. (2012). *Machine Learning: A Probabilistic Perspective*. London, UK: The MIT Press.

Nguyen, L., and Goulet, J.-A. (2018a). Anomaly detection with the switching kalman filter for structural health monitoring. *Struct. Control Health Monitor.* 25:e2136. doi: 10.1002/stc.2136

Nguyen, L., and Goulet, J.-A. (2018b). Structural health monitoring with dependence on non-harmonic periodic hidden covariates. *Eng. Struct.* 166, 187–194. doi: 10.1016/j.engstruct.2018.03.080

Prado, R., and West, M. (2010). *Time Series: Modeling, Computation, and Inference, 1st Edn*. New York, NY: Chapman & Hall/CRC.

Rao, S. S. (2008). *A Course in Time Series Analysis*. Technical report, Texas A&M University.

Salazar, F., and Toledo, M. (2018). Discussion on "thermal displacements of concrete dams: accounting for water temperature in statistical models". *Eng. Struct.* 171, 1071–1072. doi: 10.1016/j.engstruct.2015.08.001

Särkkä, S. (2013). *Bayesian Filtering and Smoothing*, Vol. 3. New York, NY: Cambridge University Press.

Scholkopf, B. and Smola, A. J. (2001). *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. Cambridge, MA: MIT Press.

Shcherbakov, M. V., Brebels, A., Shcherbakova, N. L., Tyukov, A. P., Janovsky, T. A., and Kamaev, V. A. (2013). A survey of forecast error measures. *World Appl. Sci. J.* 24, 171–176. doi: 10.5829/idosi.wasj.2013.24.itmies.80032

Simon, D. (2006). *Optimal State Estimation: Kalman, H Infinity, and Nonlinear Approaches*. New Jersey, NJ: Wiley.

Solhjell, I. (2009). *Bayesian Forecasting and Dynamic Models Applied to Strain Data from the Göta river Bridge*. Master's thesis, University of Oslo.

Tatin, M., Briffaut, M., Dufour, F., Simon, A., and Fabre, J.-P. (2015). Thermal displacements of concrete dams: accounting for water temperature in statistical models. *Eng. Struct.* 91, 26–39. doi: 10.1016/j.engstruct.2015.01.047

West, M., and Harrison, J. (1999). *Bayesian Forecasting and Dynamic Models*. New York, NY: Springer.

Westgate, R., Koo, K.-Y., and Brownjohn, J. (2015). Effect of vehicular loading on suspension bridge dynamic properties. *Struct. Infrastruct. Eng.* 11, 129–144. doi: 10.1080/15732479.2013.850731

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Nguyen, Gaudot, Khazaeli and Goulet. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

APPENDIX A

The model matrices $\{\mathbf{A}_t, \mathbf{C}_t, \mathbf{Q}_t, \mathbf{R}_t\}$ for the case-study of traffic load data are defined following

$$\mathbf{A}_t = \text{block diag}\left(1, \begin{bmatrix} 0 & \tilde{\boldsymbol{k}}_t^{\text{KR}} \\ \mathbf{0} & \mathbf{I}_{100} \end{bmatrix}, \phi^{\text{AR}}\right)$$

$$\mathbf{C}_t = [1, 1, 0_1, 0_2, \cdots, 0_{100}, 1]$$

$$\mathbf{R}_t = \left[(\sigma_v)^2\right]$$

$$\mathbf{Q}_t = \text{block diag}\left(\left(\sigma_w^{\text{B}}\right)^2, \begin{bmatrix} \left(\sigma_{w,0}^{\text{KR}}\right)^2 & \mathbf{0} \\ \mathbf{0} & \left(\sigma_{w,1}^{\text{KR}}\right)^2 \cdot \mathbf{I}_{100} \end{bmatrix}, \left(\sigma_w^{\text{AR}}\right)^2\right),$$

where $\tilde{\boldsymbol{k}}_t^{\text{KR}} = [\tilde{k}_{t,1}^{\text{KR}}, \tilde{k}_{t,2}^{\text{KR}}, \cdots, \tilde{k}_{t,100}^{\text{KR}}]$ is the normalized kernel values.

APPENDIX B

The model matrices $\{\mathbf{A}_t, \mathbf{C}_t, \mathbf{Q}_t, \mathbf{R}_t\}$ for the case-study of piezometric pressure data are defined following

$$\mathbf{A}_t = \text{block diag}\left(\begin{bmatrix} 1 & \Delta t \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & \tilde{\boldsymbol{k}}_t^{\text{KR}} \\ \mathbf{0} & \mathbf{I}_{20} \end{bmatrix}, \phi^{\text{AR}}\right)$$

$$\mathbf{C}_t = [1, 0, 1, 0_1, 0_2, \cdots, 0_{20}, 1]$$

$$\mathbf{R}_t = \left[(\sigma_v)^2\right]$$

$$\mathbf{Q}_t = \text{block diag}\left(\left(\sigma_w^{\text{B}}\right)^2 \cdot \begin{bmatrix} \frac{\Delta t^3}{3} & \frac{\Delta t^2}{2} \\ \frac{\Delta t^2}{2} & \Delta t \end{bmatrix}, \begin{bmatrix} \left(\sigma_{w,0}^{\text{KR}}\right)^2 & \mathbf{0} \\ \mathbf{0} & \left(\sigma_{w,1}^{\text{KR}}\right)^2 \cdot \mathbf{I}_{20} \end{bmatrix}, \left(\sigma_w^{\text{AR}}\right)^2\right),$$

where $\tilde{\boldsymbol{k}}_t^{\text{KR}} = [\tilde{k}_{t,1}^{\text{KR}}, \tilde{k}_{t,2}^{\text{KR}}, \cdots, \tilde{k}_{t,20}^{\text{KR}}]$ is the normalized kernel values and Δt is the time step at the time t.

APPENDIX C

The mathematical formulations for the Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) are written as following

$$\text{MAE} = \frac{1}{\text{T}} \sum_{t=1}^{\text{T}} |\hat{y}_t - y_t|$$

$$\text{RMSE} = \sqrt{\frac{\sum_{t=1}^{\text{T}} \left(\hat{y}_t - y_t\right)^2}{\text{T}}},$$

where $\hat{y}_t$ is the forecast value and y_t is the observation at time t.