Engineering Structures 166 (2018) 187–194

Contents lists available at ScienceDirect

Engineering Structures

journal homepage: www.elsevier.com/locate/engstruct

Structural Health Monitoring with dependence on non-harmonic periodic hidden covariates

L.H. Nguyen*, J-A. Goulet

Department of Civil, Geological, and Mining Engineering, Ecole Polytechnique de Montreal, 2900 Edouard Montpetit Blvd., Montreal, QC H3T 1J4, Canada

ARTICLE INFO

Keywords:
Bayesian
Dynamic Linear Models
Hidden covariates
Structural Health Monitoring
Kalman filter
Dam

ABSTRACT

In Structural Health Monitoring, non-harmonic periodic hidden covariate typically arises when an observed structural response depends on unobserved external effects such as temperature or loading. This paper addresses this challenge by proposing a new extension to Bayesian Dynamic Linear Models (BDLMs) for handling situations where non-harmonic periodic hidden covariates may influence the observed responses of structures. The potential of the new approach is illustrated on the data recorded on a dam in Canada. A model employing the proposed approach is compared to another that only uses a superposition of harmonic hidden components available from the existing BDLMs. The comparative study shows that the proposed approach succeeds in estimating hidden covariates and has a better predictive performance than the existing method using a superposition of harmonic hidden components.

1. Introduction

Structural Health Monitoring (SHM) is a key part in ensuring the long-term sustainability of our ageing structures. The SHM consists in providing the structure's health and conditions during its life service using instrumentation-based monitoring [1,2]. The measured quantities being interpreted are commonly displacements and acceleration, that is, *observed structural responses*. The important aspect in the SHM is to early detect changes in the structural behavior by interpreting the observed structural responses in order to provide infrastructure maintenance in time. As a matter of fact, the observed structural responses are commonly dependent on the environmental and operational conditions, i.e. *external effects*, such as temperature, traffic load, wind, and humidity [3–5]. In the context of SHM, an unobserved external effect is defined as a *hidden covariate*. In most cases, the hidden covariate is regrouped in two main categories: *harmonic* and *non-harmonic* hidden covariates. Fig. 1a and b present an example of a harmonic signal and of a non-harmonic signal but periodic, respectively. In the scope of this paper, we focus on the *non-harmonic periodic hidden covariates*. Non-harmonic periodic covariates are common when analyzing the behavior of structures, for example, the effect of water temperature in the field of dam engineering [6–8] or the effect of traffic load in the field of bridge engineering [9]. For the anomalous detection [10–12], a well separation of the changes due to the external effects and structural behavior is essential to reduce the *false alarms*.

The current factor limiting widespread SHM applications is the lack of generic data-interpretation methods that can be employed at low cost, for any structures. For the context of SHM applications where data is acquired periodically over a long time period, Goulet [13] proposed to address this challenge by building on the work done in the fields of Machine Learning in what is known as *State-Space Models* [14], in Applied Statistics what is known as *Bayesian Dynamic Linear Models* (BDLMs) [15–17], and in Control theory in what is known as the Kalman filter [18]. This methodology consists in employing the BDLMs to decompose the time series recorded on structures into a set of generic hidden components, each described by one or more hidden state variables. The set of available components includes, for example, a *local level component* to model the baseline response of structures, a *local trend component* to model the rate of change, a *periodic component* to model the periodic external effects, an *autoregressive component* to describe time-dependent model approximation errors, and a *regression component* to include the effect of an observed covariate on the structural response.

The BDLMs can handle harmonic covariates such as the effect of temperature on the structural response. Moreover, this can be achieved whether or not the temperature is observed. However, one limitation of BDLMs is that it is unable to handle non-harmonic periodic covariates unless they are directly observed. The requirement that non-harmonic periodic covariates must be directly observed is a difficult constraint for SHM applications where the covariates is often non-harmonic yet, observations are seldom available.

In the field of dam engineering, a common approach employed to

* Corresponding author.
E-mail addresses: luongha.nguyen@gmail.com (L.H. Nguyen), james.a.goulet@gmail.com (J.-A. Goulet).

https://doi.org/10.1016/j.engstruct.2018.03.080
Received 19 June 2017; Received in revised form 7 February 2018; Accepted 26 March 2018

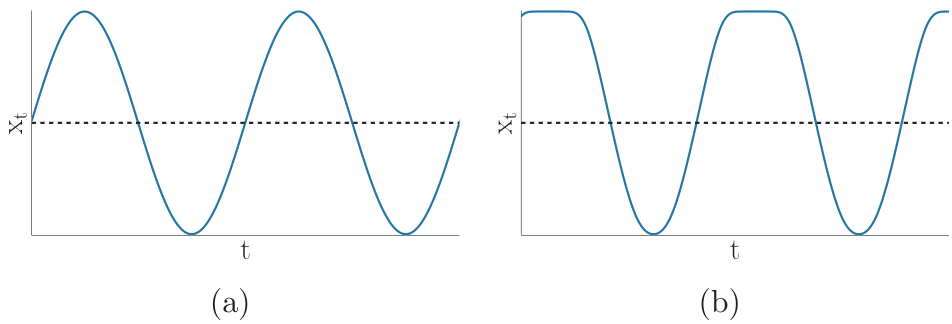

Fig. 1. The sine-like signal in (a) is harmonic and can already be handled by the BDLM method whether or not this component is observed. The signal in (b) is not harmonic. This case can only be handled by the BDLM method if the component is directly observed.

interpret SHM data is the *Hydrostatic-Seasonal-Time* (HST) method. This method has been employed in many case studies [19–22] to interpret displacement, pressure, and flow-rate observations. The main idea of HST is to separate the observations into reversible (hydrostatic and seasonal) and irreversible components. Classic HST formulations cannot handle the situation where the observations depends on non-harmonic periodic covariates [8]. Similar methods such as *Hydrostatic-Temperature-Time* (HTT) [6,23] and HST-Grad [8] employs directly the observed external effects such as concrete and water temperatures for addressing this limitation. When those data are not available, a superposition of harmonic functions can be employed for building in the non-harmonic periodic covariates [24]. The limitation is that it requires a large number of harmonic functions when it comes to the complex non-harmonic periodic covariates.

Another alternative to HST-Grad is *Neural Networks* (NN) that have shown its potential on interpreting the dam-displacement data in several applications [25–28]. NN method consists in building the function that links the displacement to time-dependent covariates such as temperature and water level by a succession of interconnected hidden layers. However, these methods are typically difficult to interpret and requires a large amount of data points. To tackle these limitations, Salazar et al. [29–31] have proposed a novel approach employed *Boosted Regression Trees* (BRTs) for analyzing the dam responses. Moreover, according to the authors the BRTs has better predictive performance than the HST and NN methods.

Although all above methods can handle non-harmonic periodic covariates using its data-recorded on the dam, they are limited in comparison with BDLM because they are based on the theory of linear regression analysis [32]. Despite having played a key historic role, linear regression is not up to the state-of-the-art approaches in the field of machine learning [14,33]. The key limitation of linear regression is that it does not distinguish between interpolating between observed data and extrapolating beyond observations. Linear regression is also known to be sensitive to outliers, prone to overfitting, and unable to handle *auto-correlation* which is omnipresent in time-series data [14].

This paper proposes a new extension to the existing BDLMs for handling situations where hidden non-harmonic periodic covariates may influence the observed responses of structures. The paper is separated into three main parts. The first part presents a summary of existing the BDLM formulation. The second part describes the approach proposed to enable the estimation of hidden non-harmonic periodic covariates. The final part illustrates the potential of the new approach on data recorded on a dam located in Canada.

2. Bayesian Dynamic Linear Models

This section presents a summary of the mathematical formulation employed by Bayesian Dynamic Linear Models (BDLMs) [13]. A BDLM is defined by its observation and transition equations which are defined as

$$\mathbf{y}_t = \mathbf{C}_t\mathbf{x}_t + \mathbf{v}_t, \quad \begin{cases} \mathbf{y}_t \sim \mathcal{N}(\mathbb{E}[\mathbf{y}_t],\mathrm{cov}[\mathbf{y}_t]) \\ \mathbf{x}_t \sim \mathcal{N}(\boldsymbol{\mu}_t,\boldsymbol{\Sigma}_t) \\ \mathbf{v}_t \sim \mathcal{N}(\mathbf{0},\mathbf{R}_t) \end{cases} \tag{1}$$

$$\mathbf{x}_t = \mathbf{A}_t\mathbf{x}_{t-1} + \mathbf{w}_t, \quad \{\mathbf{w}_t \sim \mathcal{N}(\mathbf{0},\mathbf{Q}_t). \tag{2}$$

$\mathbf{y}_t$ is the observations at the time $t \in (1:T)$ and $\mathbf{x}_t$ describes hidden state variables that they are not directly observed. Observations are modeled over time as a function of hidden state variables $\mathbf{x}_t$, an observation matrix $\mathbf{C}_t$, and a Gaussian measurement error $\mathbf{v}_t$ with mean zero and covariance matrix $\mathbf{R}_t$. The transition of hidden state variables $\mathbf{x}_t$ between time steps are defined by the transition matrix $\mathbf{A}_t$ and a Gaussian model error $\mathbf{w}_t$ with mean zero and covariance matrix $\mathbf{Q}_t$. The main strength of BDLMs for SHM applications is the capacity to model a variety number of structural responses from a limited set of generic hidden components such as basis levels, local trends, periodic components and regression components. See Goulet [13] and West & Harrison [17] for the full description of generic hidden components. In BDLMs, the hidden state variables $\mathbf{x}_t$ at a time t are estimated using observations $\mathbf{y}_{1:t}$ and the *Kalman filter* (KF) algorithm. This algorithm is an iterative two-steps mathematical process that estimates the posterior mean vector $\boldsymbol{\mu}_{t|t}$ and covariance matrix $\boldsymbol{\Sigma}_{t|t}$ so that

Prediction step

$$\begin{array}{rcll} p(\mathbf{x}_t|\mathbf{y}_{1:t-1}) & = & \mathcal{N}(\mathbf{x}_t;\boldsymbol{\mu}_{t|t-1},\boldsymbol{\Sigma}_{t|t-1}) & \text{Prior state estimate} \\ \boldsymbol{\mu}_{t|t-1} & \triangleq & \mathbf{A}_t\boldsymbol{\mu}_{t-1|t-1} & \text{Prior expected value} \\ \boldsymbol{\Sigma}_{t|t-1} & \triangleq & \mathbf{A}_t\boldsymbol{\Sigma}_{t-1|t-1}\mathbf{A}_t^{\intercal} + \mathbf{Q}_t & \text{Prior covariance} \end{array}$$

Measurement step

$$\begin{array}{rcll} p(\mathbf{x}_t|\mathbf{y}_{1:t}) & = & \mathcal{N}(\mathbf{x}_t;\boldsymbol{\mu}_{t|t},\boldsymbol{\Sigma}_{t|t}) & \text{Posterior state estimate} \\ \boldsymbol{\mu}_{t|t} & = & \boldsymbol{\mu}_{t|t-1} + \mathbf{K}_t\mathbf{r}_t & \text{Posterior expected value} \\ \boldsymbol{\Sigma}_{t|t} & = & (\mathbf{I}-\mathbf{K}_t\mathbf{C}_t)\boldsymbol{\Sigma}_{t|t-1} & \text{Posterior covariance} \\ \mathbf{r}_t & \triangleq & \mathbf{y}_t-\hat{\mathbf{y}}_t & \text{Innovation vector} \\ \hat{\mathbf{y}}_t & \triangleq & \mathbb{E}[\mathbf{y}_t|\mathbf{y}_{1:t-1}] = \mathbf{C}_t\boldsymbol{\mu}_{t|t-1} & \text{Predicted observations vector} \\ \mathbf{K}_t & \triangleq & \boldsymbol{\Sigma}_{t|t-1}\mathbf{C}_t^{\intercal}\mathbf{G}_t^{-1} & \text{Kalman gain matrix} \\ \mathbf{G}_t & \triangleq & \mathbf{C}_t\boldsymbol{\Sigma}_{t|t-1}\mathbf{C}_t^{\intercal} + \mathbf{R}_t & \text{Innovation covariance matrix.} \end{array}$$

The Kalman filter algorithm uses the *Kalman gain* $\mathbf{K}_t$ to weight the information coming from observations $\mathbf{y}_t$, in comparison with the information coming from prior knowledge.

The model matrices $\{\mathbf{A}_t,\mathbf{C}_t,\mathbf{Q}_t,\mathbf{R}_t\}$ contain several parameters $\mathcal{P}$ that need to be estimated. A common approach for this task is to employ *Maximum Likelihood Estimation* (MLE). Maximum likelihood estimates are obtained by maximizing the joint prior probability of observations with the hypothesis that observations $\mathbf{y}_{1:T}$ are independent of each other so that

$$p(\mathbf{y}_{1:T}|\mathcal{P}) = \prod_{t=1}^{T} p(\mathbf{y}_t|\mathbf{y}_{1:t-1},\mathcal{P}). \tag{3}$$

For the purpose of improving the numerical stability, one can sum the