Structural Health Monitoring with Dependence on Non-Harmonic Periodic Hidden Covariates

L.H. Nguyen[a,*], J-A. Goulet[a]

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

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.

Keywords: Bayesian, Dynamic Linear Models, Hidden Covariates, Structural Health Monitoring, Kalman Filter, Dam.

*Corresponding author

Email addresses: luongha.nguyen@gmail.com (L.H. Nguyen), james.a.goulet@gmail.com (J-A. Goulet)

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, 4, 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, 7, 8] or the effect of traffic load in the field of bridge engineering [9]. For the anomalous detection [10, 11, 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

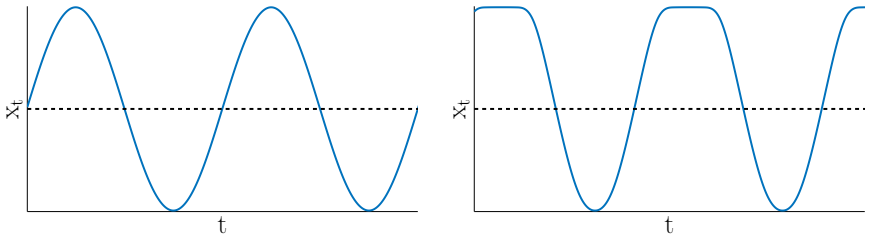


Fig. 1. The sine-like signal in (a) is harmonic and can already be handled by the BDLM method whether or not this component is observed. The signal in (b) is not harmonic. This case can only be handled by the BDLM method if the component is directly observed.

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, 16, 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 temper-

ature 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 interpret SHM data is the *Hydrostatic-Seasonal-Time* (HST) method. This method has been employed in many case studies [19, 20, 21, 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, 26, 27, 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, 30, 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

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

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

$\mathbf{y}_t$ 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

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

Measurement step

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

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 natural logarithm of the marginal prior probability of observations so that Eq. (3) is rewritten as

$$\begin{aligned} \ln p(\mathbf{y}_{1:T}|\mathcal{P}) &= \sum_{t=1}^{T} \ln p(\mathbf{y}_t|\mathbf{y}_{1:t-1},\mathcal{P}) \\ &= \sum_{t=1}^{T} \ln \left[\mathcal{N}(\mathbf{y}_t; \mathbf{C}_t\boldsymbol{\mu}_{t|t-1}, \mathbf{R}_t + \mathbf{C}_t\boldsymbol{\Sigma}_{t|t-1}\mathbf{C}_t^{\intercal})\right] \end{aligned} \tag{4}$$

Eq. (4) is called the *log-likelihood function*. Optimal parameters $\mathcal{P}^*$ are identified by maximizing Eq. (4) using convex optimization algorithms. The maximization approach employed in this paper is the *Newton-Raphson* algorithm [32].

3. Methodology for Estimating Hidden Non-harmonic Covariates

When observed, covariates affecting the responses of structures, e.g. temperature or loading can be included as a regressor in a Bayesian Dynamic Linear Model (BDLM). One approach to include the effect of these observed covariates is to employ a *Dynamic Regression Component* (DRC) [17]. In the DRC, the dynamic regression coefficient is treated as an unknown state variable $x_t^{\mathtt{DR}}$, whose temporal evolution follows a random walk. This random walk is parameterized by a transition matrix $\mathbf{A}_t^{\mathtt{DR}} = 1$, and by a transition covariance matrix $\mathbf{Q}_t^{\mathtt{DR}} = (\sigma^{\mathtt{DR}})^2$. For $\sigma^{\mathtt{DR}} = 0$, the dynamic regression coefficient $x_t^{\mathtt{DR}}$ is assumed to be stationary in time, for $\sigma^{\mathtt{DR}} > 0$, the dynamic regression coefficient $x_t^{\mathtt{DR}}$ is assumed to be changing over time (non-stationary). The regressor, i.e. the observed covariate $y_t^{\mathtt{DR}}$, is placed directly in the observation matrix so that $\mathbf{C}_t^{\mathtt{DR}} = y_t^{\mathtt{DR}}$.

The new methodology proposes to build on the dynamic regression component formulation in order to provide a method capable of handling hidden, yet non-harmonic covariates. For this case, the block component matrices are

$$\mathbf{A}_t^{\mathtt{DR}} = 1,\ \mathbf{C}_t^{\mathtt{DR}} = h(t, \mathcal{D}),\ \mathbf{Q}_t^{\mathtt{DR}} = (\sigma^{\mathtt{DR}})^2.$$

In this new formulation, observation matrix $\mathbf{C}_t^{\mathtt{DR}}$ is replaced by the hidden response function $h(\mathcal{D}, t)$. The function $h(\mathcal{D}, t)$ consists in a cubic spline [34] capable of interpolating hidden covariate values at any time stamps. Fig. 2 presents an example of a *hidden response* function $h(\mathcal{D}, t)$. $h(\mathcal{D}, t)$ depends

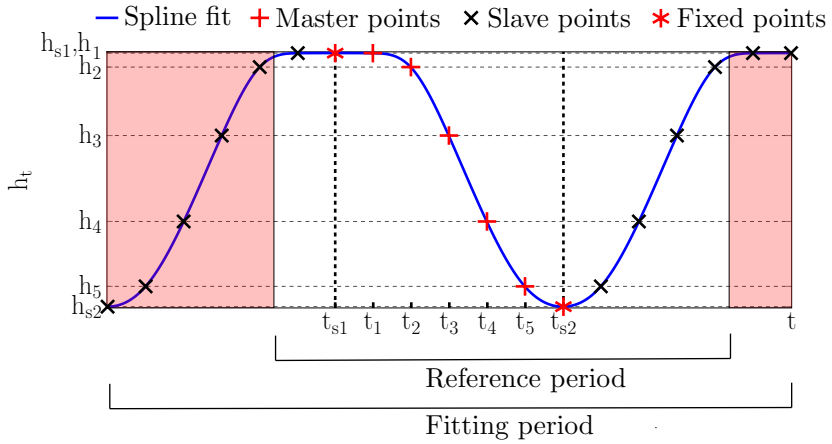


Fig. 2. Example spline fitted using master and slave control points defined over three sub-segments that are separated by the vertical symmetry lines.

on the time t as well as on a set of D *master control points* $\mathcal{D} = \{(t_i, h_i), \forall i = 1 : D\}$, where $h_i \in [-1, 1]$ is the *normalized hidden covariate value* (NHCV) and t_i is the time stamp corresponding to h_i. Note that the amplitude of the hidden covariates that influence on the structural responses, is defined not by $h(\mathcal{D}, t)$ but by the dynamic regression coefficient $x_t^{\mathtt{DR}}$. The methodology can take advantage of the periodicity of the studied phenomenon in order

to identify only (t_i, h_i) for master control points that are defined over the domain (1) having a duration corresponding to half a period, and (2) bounded at each end by symmetry planes. Note that if no symmetry planes exist, the same method applies except that the number of control points increases. The NHCVs h_{s1} and h_{s2} for time stamps corresponding to the symmetry planes, t_{s1} and t_{s2}, are fixed at either -1 or 1. Time stamps t_i for the master control points are uniformly spaced between t_{s1} and t_{s2}. Over one half-period before and after the symmetry planes, *slave control points* are defined in order to constrain the spline slope for the *fixed points* (t_{s1}, h_{s1}) and (t_{s2}, h_{s2}). Slave control points are replicates of the master points defined using the symmetry condition with respect to either t_{s1} or t_{s2}. Although the spline is fitted over the *fitting period* including the entire set of slave and master control points, only a portion having a length of one period is employed. This sub-selection is called *reference period*. An example of spline fitted using a set of five control points $\mathcal{D} = \{(t_i, h_i), \forall i = 1:5\}$ is presented in Fig. 2. Master control points are represented by plus signs, slave points by crosses, and fixed points by asterisks. Vertical dashed lines represent symmetry planes with respect to time. In this example, there are five NHCVs $\{h_1, \cdots, h_5\}$ that needs to be estimated from data. Once NHCVs have been estimated using data, the hidden response $h(\mathcal{D}, t)$ is generalized for any time stamps t by extracting the spline value corresponding to any day of the year within the reference period.

In practical applications, the NHCVs h_i are unknown and need to be estimated indirectly from observations of a structure's behavior. For that purpose, these NHCVs are added to the set of parameters $\mathcal{P}$ to be estimated by maximizing the log-likelihood function in Eq. (4).

4. Case-Study: A Dam in Canada

The potential of the new approach for handling the hidden non-harmonic covariates is illustrated on the displacement data recorded on gravity dam located in Canada. Fig. 3 shows the location of the sensor which is found on the west bank of the dam. This sensor employs an inverted pendulum system for monitoring the displacements of the gravity dam along three orthogonal directions. The X-direction points toward the West Bank, the Y-direction follows the water flow and the Z-direction points upward.

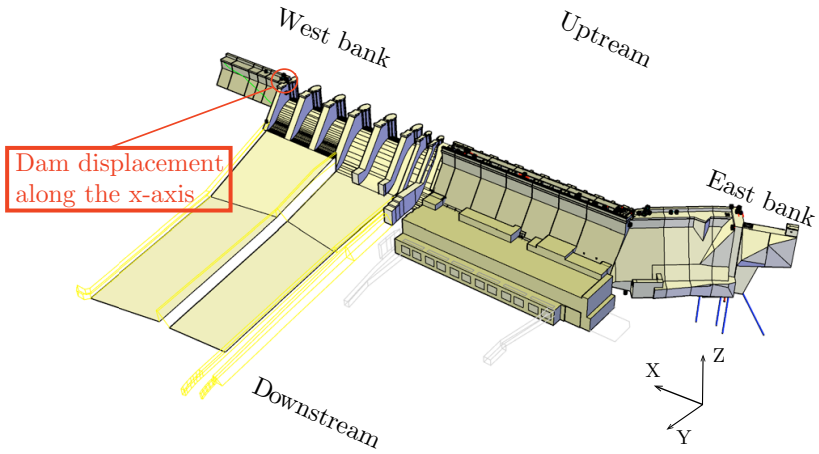


Fig. 3. Location plan of sensors deployed across the structure to monitor the dam's behavior.

4.1. Data Description

This paper studies the horizontal dam displacement data along the X-direction measured using an inverted pendulum over a period of five years. Engineers responsible for the dam instrumentation have estimated the observation error standard deviation to be approximately 0.3 mm. The complete

dataset is presented in Fig. 4. In addition to the linear trend, the data displays

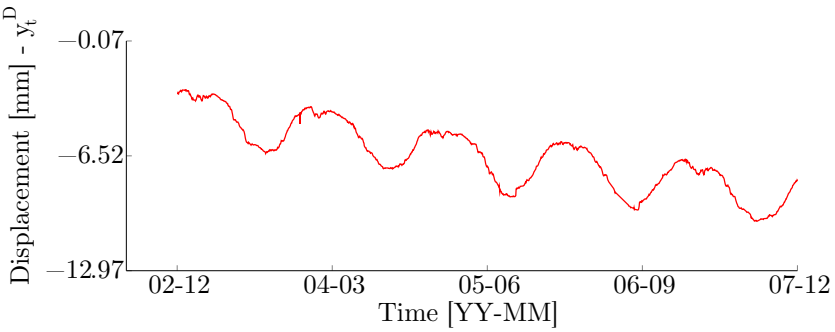


Fig. 4. The X-direction displacement collected over the period of five years.

a yearly seasonal pattern where displacements are maximal during winter months and minimal during summer. The key aspect here is that the seasonal effect is non-harmonic; the evolution of displacement during the winter is slower than during the summer. The hypothesis for this behavior is that the structure's response depends not only on the air temperature but also on the water temperature that is known to follow a non-harmonic pattern where in winter months, the temperature stabilizes despite the air temperature dropping below -20°C. The issue here is that no data is available from 2002 to 2007 in order to employ water temperature as regressor in BDLM.

In addition to the hidden non-harmonic covariate, another challenge is the non-uniformity of time-step recordings. Fig. 5 presents the time-step variation for the entire dataset duration. This challenge is addressed by defining parameters as a function of the time-step length. For that purpose, the reference time step length is defined as 24 hours.

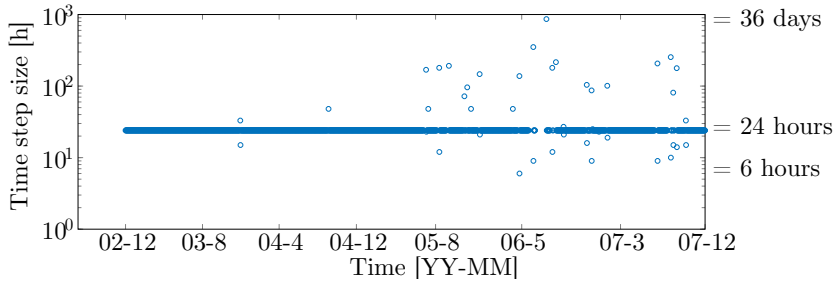


Fig. 5. Time-step size is presented in a log space for the dataset duration. Note that most of displacement data are recorded for a period of a 24 hours, which time-steps can vary in the range between 6 hours to 36 days.

4.2. Hidden Covariate Construction

From the raw displacement-data presented Fig. 4, it is estimated that the reference period of hidden covariates is 365 days and that the two fixed points corresponding to symmetry planes are located at the 50^{th} end 232.5^{th} day of each year. A set of 5 control points $\mathcal{D} = \{(t_i, h_i), \forall i = 1:5\}$ is defined where t_i are uniformly spaced within the interval $(50, 232.5)$. The slave points are defined using the symmetry planes. Fig. 6 presents the hidden component constructed for the entire dataset using spline values defined over the reference period. In this figure, the position of normalized hidden covariate values $\{h_1, \cdots, h_5\}$ are estimated using the training data.

4.3. Model Comparison

This section examines the performance of the new approach for handling the hidden covariates. For this purpose, a model employing the new approach is compared to another model that uses a superposition of harmonic hidden components available from the existing BDLMs. The 1^{st} model employing the proposed method is called *model-DR*, where the DR stands for dynamic

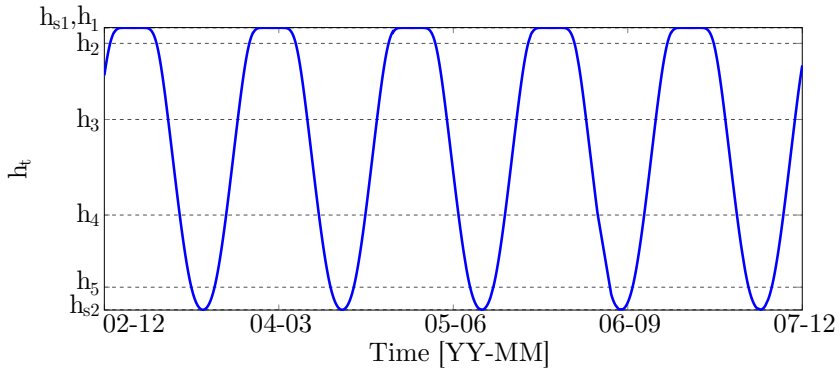


Fig. 6. The hidden component for the entire dataset

regression. In the model-DR, the observations are decomposed into a vector of four hidden components : *a local level, a local trend, a dynamic regression component, and an autoregressive component.* The 2^{nd} model is denoted by *model-S*, where S stands for superposition. The vector of hidden components for the model-S is the same as for the model-DR, except the dynamic regression component is replaced by two harmonic hidden components with a period of 365 and 180 days. Here, the vectors of hidden state variables corresponding to model-DR and model-S are respectively,

$$\begin{aligned}
\text{DR}: \begin{cases} \mathbf{x}_t = [\underbrace{x_t^{\texttt{D,LL}}}_{\text{local level}} , \underbrace{x_t^{\texttt{D,LT}}}_{\text{local trend}} , \underbrace{x_t^{\texttt{D,DR}}}_{\text{dynamic regression}} , \underbrace{x_t^{\texttt{D,AR}}}_{\text{autoregressive}}]^{\intercal} \end{cases} \\
\text{S}: \begin{cases} \mathbf{x}_t = [\underbrace{x_t^{\texttt{D,LL}}}_{\text{local level}} , \underbrace{x_t^{\texttt{D,LT}}}_{\text{local trend}} , \underbrace{x^{\texttt{T1,S1}}, x^{\texttt{T1,S2}}}_{\text{cycle, p=365 days}} , \underbrace{x^{\texttt{T2,S1}}, x^{\texttt{T2,S2}}}_{\text{cycle, p=180 days}} , \underbrace{x_t^{\texttt{D,AR}}}_{\text{autoregressive}}]^{\intercal}. \end{cases}
\end{aligned} \tag{5}$$

Complete model matrices are presented in Appendix A.

4.3.1. Parameter calibration

The unknown parameters to be estimated for model-DR and model-S are regrouped in two following sets:

$$\begin{aligned} \text{DR} &: \begin{cases} \mathcal{P} = \{\phi^{\mathtt{D,AR}}, \sigma^{\mathtt{D,LT}}, \sigma^{\mathtt{D,AR}}, h_1, h_2, h_3, h_4, h_5\} \end{cases} \\ \text{S} &: \begin{cases} \mathcal{P} = \{\phi^{\mathtt{D,AR}}, \sigma^{\mathtt{D,LT}}, \sigma^{\mathtt{D,AR}}\}, \end{cases} \end{aligned} \tag{6}$$

where the *autocorrelation coefficient*; $\phi^{\mathtt{D,AR}} \in (0, 1)$, the *normalized hidden covariate values*; $h_i \in [-1, 1]\ \forall i = 1 : 5$, the *local trend standard deviation*; $\sigma^{\mathtt{D,LT}} \in \mathbb{R}^+$, and the *autocorrelation standard deviation*; $\sigma^{\mathtt{D,AR}} \in \mathbb{R}^+$. Based on the available information about the dam instrumentation, the observation error standard deviation is fixed to $\sigma^{\mathtt{D,R}} = 0.3\,\text{mm}$. Note that the displacement data are collected over a period of five years for a total of approximately 1700 data points. The unknown model parameters are estimated using a training period of four years (1359 data points). Initial values for all state variables have been defined using engineering heuristics. Hence the means and covariances of state variables at the time $t = 0$ are

$$\begin{aligned} \text{DR} &: \begin{cases} \boldsymbol{\mu}_0 = [-4, -2.8 \times 10^{-3}, -1.3, 0] \\ \boldsymbol{\Sigma}_0 = \text{diag}([0.25, 10^{-8}, 4, 0.02508]) \end{cases} \\ \text{S} &: \begin{cases} \boldsymbol{\mu}_0 = [-4, -2.8 \times 10^{-3}, -1, -1, -1, -1, 0] \\ \boldsymbol{\Sigma}_0 = \text{diag}([0.25, 10^{-8}, 25, 25, 25, 25, 0.02508]), \end{cases} \end{aligned}$$

where the ordering of initial states is the same as Eq. (5). Note that the initial state of the AR process is defined as $\boldsymbol{\mu}_0^{\mathtt{D,AR}} = 0$, $\boldsymbol{\Sigma}_0^{\mathtt{D,AR}} = \frac{\sigma^{\mathtt{D,AR}}}{(1-\phi^{\mathtt{D,AR}})^2}$ according to its stationary property [35]. The initial parameter values are estimated using

engineering heuristics

$$\text{DR}: \begin{cases} \mathcal{P}^0 = \{0.991, 6.19 \times 10^{-7}, 0.0212, -0.95, -0.8, -0.33, 0.3, 0.7\} \end{cases}$$

$$\text{S}: \begin{cases} \mathcal{P}^0 = \{0.991, 6.19 \times 10^{-7}, 0.0212\}. \end{cases}$$

The convergence of the optimization procedure is reached when the log-likelihood between two consecutive loops satisfies

$$\begin{cases} \text{log-likelihood}_{i-1} & < & \text{log-likelihood}_i \\ \left|\text{log-likelihood}_i - \text{log-likelihood}_{i-1}\right| & \leq & 10^{-7} \times \left|\text{log-likelihood}_{i-1}\right| \end{cases},$$

where i corresponds to i^{th} optimization loop. The optimal parameter values found for the two models are

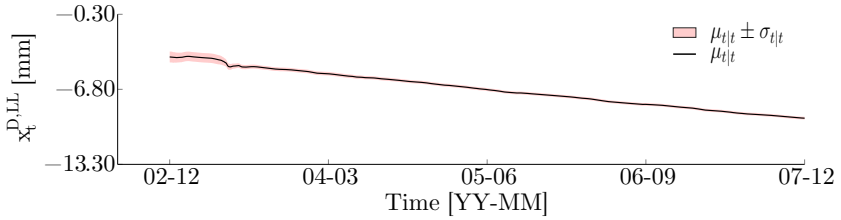
$$\text{DR}: \begin{cases} \mathcal{P}^* = \{0.993, 1.304 \times 10^{-7}, 0.028, -0.987, -0.898, -0.435, 0.286, 0.790\} \end{cases}$$

$$\text{S}: \begin{cases} \mathcal{P}^* = \{0.989, 7.664 \times 10^{-8}, 0.029\}, \end{cases}$$

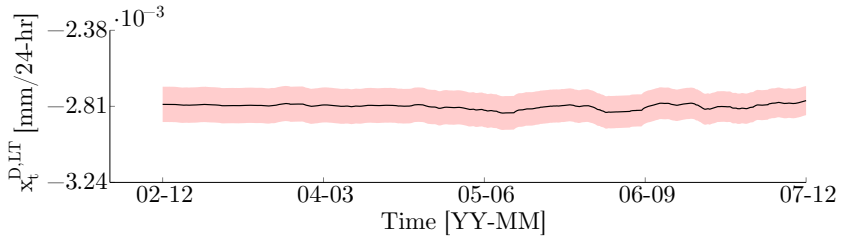
where the ordering of each parameter remains identical as in Eq. (6). The log-likelihood values of model-DR and model-S in the training period are respectively 252.2 and 236.9.

4.3.2. Results and discussion

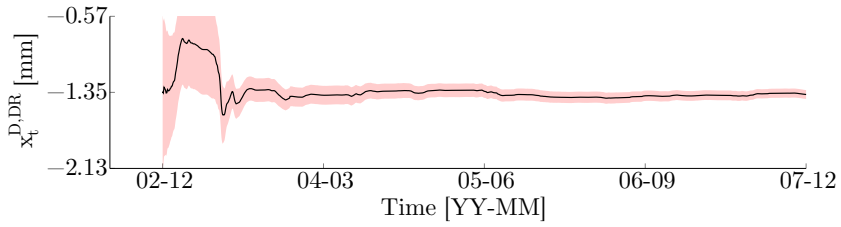
The BDLM framework employs the filtering algorithm and the set of parameters obtained in the training period to separate the observations into its hidden components. Fig. 7 and Fig. 8 present the hidden components for the entire dataset for model-DR and model-S, respectively. The solid black line represents the mean values μ and its $\pm\sigma$ standard deviation interval is represented by the shaded region. It is noticed that the local levels in



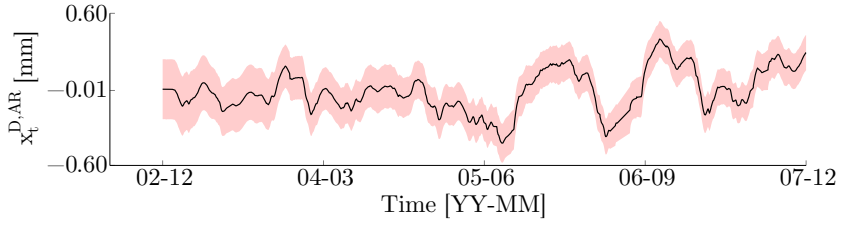
(a) Local level $x_t^{\mathtt{D,LL}}$



(b) Local trend $x_t^{\mathtt{D,LT}}$

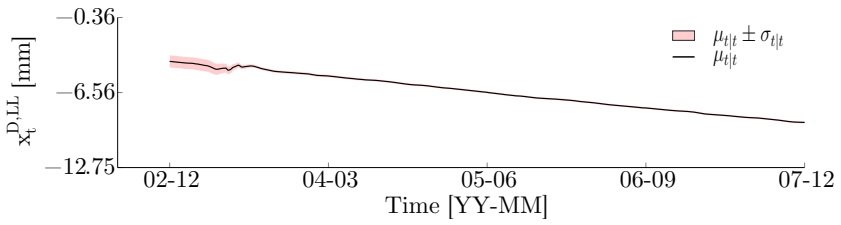


(c) Dynamic regression component $x_t^{\mathtt{D,DR}}$

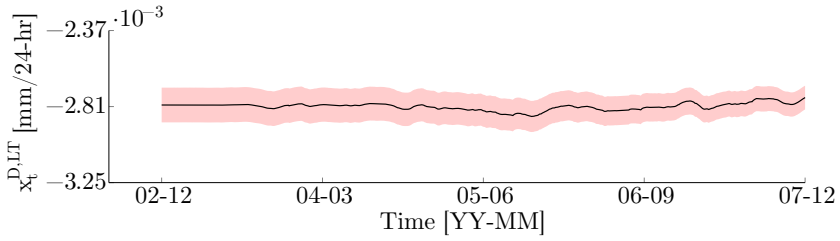


(d) Autoregresive component $x_t^{\mathtt{D,AR}}$

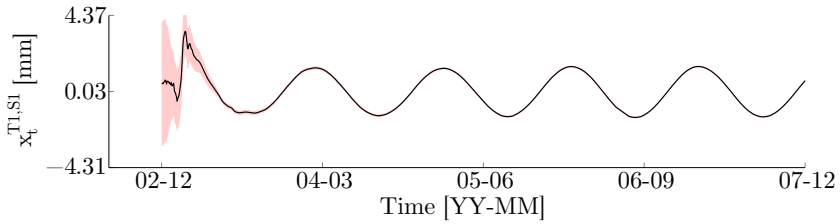
Fig. 7. Expected values $\mu_{t|t}$ and uncertainty bound $\mu_{t|t} \pm \sigma_{t|t}$ for hidden components of the model-DR.



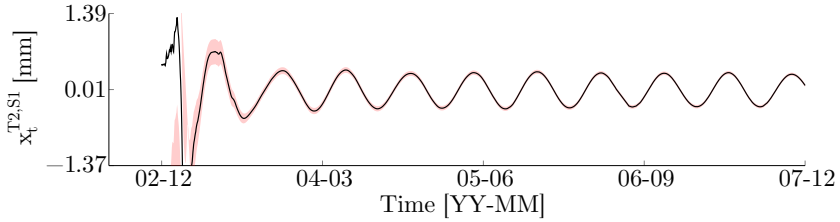
(a) Local level $x_t^{\mathtt{D,LL}}$, [mm]



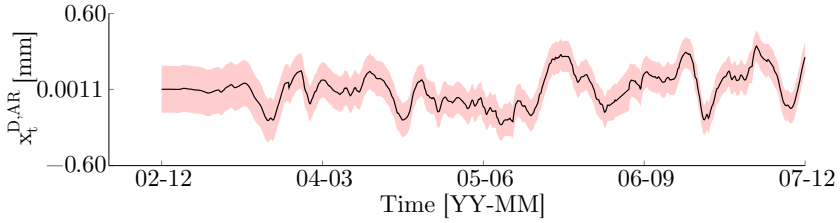
(b) Local trend $x_t^{\mathtt{D,LT}}$



(c) Cycle of 365 days $x_t^{\mathtt{T1,S1}}$



(d) Cycle of 180 days $x_t^{\mathtt{T2,S1}}$



(e) Autoregresive component $x_t^{\mathtt{D,AR}}$

Fig. 8. Expected values $\mu_{t|t}$ and uncertainty bound $\mu_{t|t} \pm \sigma_{t|t}$ for hidden components of the model-S.

Fig. 7(a) and Fig. 8(a) show a drift where the rate of change is approximately to $-1.03\,\mathrm{mm/year}$. The autoregresive components presented in Fig. 7(d) and Fig. 8(e) are stationary as expected because the autocorrelation coefficient has been constrained to the interval $(0, 1)$. The dynamic regression component shown in Fig. 7(c) is nearly constant over time. It means that the dynamic regression coefficient $x_t^{\mathtt{DR}}$ defined in Section 3 is unchanged over time. Note that the standard deviation for state variables decreases over time because the effect of imperfect initial condition vanishes as more and more data are observed.

The predictive performance of models are compared on the basis of their log-likelihood value obtained using a test set. Note that the data in the test set has not been employed to calibrate the parameters $\mathcal{P}$. Table 1 presents the log-likelihood estimates for the training and the test set for model-DR and model-S. The model-DR has a log-likelihood value for the test set that is 21.1%

Table 1. Comparison of log-likelihood estimates for the model-DR using the novel approach proposed in this paper and the model-S that uses a superposition of harmonic components.

	Log-likelihood, $\ln p(\mathbf{y}_{1:T}\|\mathcal{P}^*)$	
Model	Calibration set	Test set
	Years 1–4	5^{th} year
model-DR	252.3	64.3
model-S	236.9	53.1

greater than model-S while the computational time for the Kalman filter for

model-DR is 2.9 sec versus 1.2 sec for model-S. This confirms the superior predictive capacity of the model-DR which employs the methodology proposed in this paper. This justifies that the model-DR outperforms the model-S in separating the effect of the environmental conditions on the structural behavior. This improvement is expected to help with anomaly detection, where the distinguishing the changes caused by the environmental conditions and the structural behavior is a key part for reducing the false alarms. Note that although the computational time for the model-DR is approximately 2.5 times slower than the model-S, the computational time required for five years of dataset is negligible in comparison with the sampling period (= 24 hours).

The comparative study shows the potential of the proposed method for handling the non-harmonic periodic covariate that influence on the the observed structural response. One current limitation is that the reference period presented in Section 3, needs to be pre-defined for an effective performance. The treatment of the reference period as an unknown parameter to be inferred from data will be a subject for the future research.

4.4. Number of Master Control Points

The effect of the control point on the predictive capacity is studied by generating three models with 3, 5, and 7 control points. Table 2 shows that the relative change of the log-likelihood values for the test set with 3 and 7 control points compared to 5 control points are 1.4 % and 0.3 %, respectively. Note that the relative change is evaluated using the following formulation

$$\mathrm{RC}_p = \frac{\left|\text{log-likelihood}_p - \text{log-likelihood}_5\right|}{\left|\text{log-likelihood}_5\right|} \times 100,$$

Table 2. Comparison of log-likelihood estimates using different number of control points in the model-DR.

	Log-likelihood, $\ln p(\mathbf{y}_{1:T}\|\mathcal{P}^*)$		
Number of control points	Calibration set Years $1-4$	Test set 5^{th} year	Relative change
3	250.6	63.4	1.4 %
5	252.3	64.3	–
7	253.0	64.1	0.3 %

where log-likelihood$_p$ refers to the log-likelihood value of p control points. The difference between model-DR with $3, 7$, and 5 control points is negligible. In this case, the choice of using 5 control points is the most suited because its predictive capacity is slightly greater than the remaining control points and there is a computational gain in using 5 control points (5.87 mins) instead of 7 control points (11.60 mins). In practice, the different sets of the control point should be tested during the model development in order to ensure a reliable estimation.

5. Conclusion

This paper proposes a new extension to Bayesian Dynamic Linear Models (BDLMs) for handling situations where hidden covariates influence the observed responses of structures. The application of the new formulation

to the data recorded on a full-scale dam shows that it is able to estimate non-harmonic periodic hidden covariates. Its predictive performance is better than the existing method using a superposition of harmonic hidden components. Moreover, the computational time is negligible in comparison with the sampling period. Because non-harmonic periodic hidden covariates are common in the SHM application, this new extension to BDLMs opens the way for new practical applications.

6. Appendix A

The *transition matrix* ($\mathbf{A}_t$), the *observation matrix* ($\mathbf{C}_t$), the *observation error covariance matrix* ($\mathbf{R}_t$), and the *model error covariance matrix* ($\mathbf{Q}_t$) for model-DR and model-S are defined following

Model-DR

$$
\begin{aligned}
\mathbf{A}_t &= \text{block diag}\left(\begin{bmatrix} 1 & \Delta t \\ 0 & 1 \end{bmatrix}, 1, \phi^{\mathtt{D,AR}}\right) \\
\mathbf{C}_t &= [1, 0, h(t, \mathcal{D}), 1] \\
\mathbf{R}_t &= \left[(\sigma^{\mathtt{D,R}})^2\right] \\
\mathbf{Q}_t &= \text{block diag}\left((\sigma^{\mathtt{D,LT}})^2 \begin{bmatrix} \frac{\Delta t^3}{3} & \frac{\Delta t^2}{2} \\ \frac{\Delta t^2}{2} & \Delta t \end{bmatrix}, 0, (\sigma^{\mathtt{D,AR}})^2\right)
\end{aligned}
$$

Model-S

$$\mathbf{A}_t = \text{block diag}\left(\begin{bmatrix} 1 & \Delta t \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} \cos\omega^{\mathtt{T1}} & \sin\omega^{\mathtt{T1}} \\ -\sin\omega^{\mathtt{T1}} & \cos\omega^{\mathtt{T1}} \end{bmatrix}, \begin{bmatrix} \cos\omega^{\mathtt{T2}} & \sin\omega^{\mathtt{T2}} \\ -\sin\omega^{\mathtt{T2}} & \cos\omega^{\mathtt{T2}} \end{bmatrix}, \phi^{\mathtt{D,AR}}\right)$$

$$\mathbf{C}_t = [1, 0, 1, 0, 1, 0, 1]$$

$$\mathbf{R}_t = \left[\left(\sigma^{\mathtt{D,R}}\right)^2\right]$$

$$\mathbf{Q}_t = \text{block diag}\left(\left(\sigma^{\mathtt{D,LT}}\right)^2 \begin{bmatrix} \frac{\Delta t^3}{3} & \frac{\Delta t^2}{2} \\ \frac{\Delta t^2}{2} & \Delta t \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, \left(\sigma^{\mathtt{D,AR}}\right)^2\right),$$

where Δt is the time step at the time t.

Acknowledgements

This project is funded by the Natural Sciences and Engineering Research Council of Canada (NSERC, RGPIN-2016-06405). The authors acknowledge the contribution of Hydro-Quebec who provided the dataset employed in this research. More specifically, the authors thank Benjamin Miquel and Patrice Côté from Hydro-Quebec for their help in the project.

References

[1] P. Cawley, R. Adams, The location of defects in structures from measurements of natural frequencies, The Journal of Strain Analysis for Engineering Design 14 (2) (1979) 49–57.

[2] R. Adams, P. Cawley, C. Pye, B. Stone, A vibration technique for non-destructively assessing the integrity of structures, Journal of Mechanical Engineering Science 20 (2) (1978) 93–100.

[3] Y. Ni, X. Hua, K. Fan, J. Ko, Correlating modal properties with temperature using long-term monitoring data and support vector machine technique, Engineering Structures 27 (12) (2005) 1762 – 1773. doi:https://doi.org/10.1016/j.engstruct.2005.02.020.

[4] K.-V. Yuen, S.-C. Kuok, Ambient interference in long-term monitoring of buildings, Engineering Structures 32 (8) (2010) 2379 – 2386. doi:https://doi.org/10.1016/j.engstruct.2010.04.012.

[5] K.-V. Yuen, S.-C. Kuok, Modeling of environmental influence in structural health assessment for reinforced concrete buildings, Earthquake Engineering and Engineering Vibration 9 (2) (2010) 295–306. doi:10.1007/s11803-010-0014-4.

[6] P. Léger, M. Leclerc, Hydrostatic, temperature, time-displacement model for concrete dams, Journal of engineering mechanics 133 (3) (2007) 267–277. doi:10.1061/(ASCE)0733-9399(2007)133:3(267).

[7] F. Salazar, M. Toledo, Discussion on "thermal displacements of concrete dams: Accounting for water temperature in statistical models", Engineering Structuresdoi:10.1016/j.engstruct.2015.08.001.

[8] M. Tatin, M. Briffaut, F. Dufour, A. Simon, J.-P. Fabre, Thermal displacements of concrete dams: Accounting for water temperature

in statistical models, Engineering Structures 91 (2015) 26 – 39. doi:10.1016/j.engstruct.2015.01.047.

[9] J.-A. Goulet, K. Koo, Empirical validation of bayesian dynamic linear models in the context of structural health monitoring, Journal of Bridge Engineering 23 (2) (2018) 05017017. doi:10.1061/(ASCE)BE.1943-5592.0001190.

[10] M. D. Spiridonakos, E. N. Chatzi, B. Sudret, Polynomial chaos expansion models for the monitoring of structures under operational variability, ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering 2 (3) (2016) B4016003.

[11] S. Bogoevska, M. Spiridonakos, E. Chatzi, E. Dumova-Jovanoska, R. Höffer, A novel bi-component structural health monitoring strategy for deriving global models of operational wind turbines, in: Proceedings of the 8th European Workshop on Structural Health Monitoring (EWSHM 2016), Bilbao, Spain, 2016, pp. 5–8.

[12] L. H. Nguyen, J.-A. Goulet, Anomaly detection with the switching kalman filter for structural health monitoring, Structural Control and Health Monitoring (2018) e2136–n/adoi:10.1002/stc.2136.

[13] J.-A. Goulet, Bayesian dynamic linear models for structural health monitoring, Structural Control and Health Monitoring 24 (2017) e2035–n/a. doi:10.1002/stc.2035.

[14] K. P. Murphy, Machine learning: a probabilistic perspective, The MIT Press, 2012.

[15] I. Solhjell, Bayesian forecasting and dynamic models applied to strain data from the göta river bridge, Master's thesis, University of Oslo, Oslo (2009).

[16] M. West, Bayesian dynamic modelling, Bayesian Inference and Markov Chain Monte Carlo: In Honour of Adrian FM Smith (2013) 145–166.

[17] M. West, J. Harrison, Bayesian Forecasting and Dynamic Models, Springer Series in Statistics, Springer New York, 1999.

[18] B. P. Gibbs, Advanced Kalman filtering, least-squares and modeling: a practical handbook, John Wiley & Sons, 2011.

[19] S. Ferry, G. Willm, Méthodes d'analyse et de surveillance des déplacements observés par le moyen de pendules dans les barrages, in: VIth International Congress on Large Dams, 1958, pp. 1179–1201.

[20] P. Léger, S. Seydou, Seasonal thermal displacements of gravity dams located in northern regions, Journal of performance of constructed facilities 23 (3) (2009) 166–174. doi:10.1061/(ASCE)0887-3828(2009)23:3(166).

[21] F. Lugiez, N. Beaujoint, X. Hardy, L'auscultation des barrages en exploitation au service de la production hydraulique d'électricité de france, des principes aux résultats, in: Xth International Congress on Large Dams, 1970, pp. 577–600.

[22] G. Willm, N. Beaujoint, Les méthodes de surveillance des barrages au service de la production hydraulique d'electricité de france, problèmes

anciens et solutions nouvelles, in: IXth International Congress on Large Dams, 1967, pp. 529–550.

[23] J. Mata, A. Tavares de Castro, J. Sá da Costa, Constructing statistical models for arch dam deformation, Structural Control and Health Monitoring 21 (3) (2014) 423–437. doi:10.1002/stc.1575.

[24] S. Gamse, M. Oberguggenberger, Assessment of long-term coordinate time series using hydrostatic-season-time model for rock-fill embankment dam, Structural Control and Health Monitoring 24 (1) (2017) e1859–n/a, e1859 STC-15-0220.R2. doi:10.1002/stc.1859.

[25] S. Demirkaya, Deformation analysis of an arch dam using anfis, in: Proceedings of the second international workshop on application of artificial intelligence and innovations in engineering geodesy. Braunschweig, Germany, 2010, pp. 21–31.

[26] F. Kang, J. Liu, J. Li, S. Li, Concrete dam deformation prediction model for health monitoring based on extreme learning machine, Structural Control and Health Monitoring (2017) e1997–n/aE1997 STC-16-0071.R1. doi:10.1002/stc.1997.

[27] J. Mata, Interpretation of concrete dam behaviour with artificial neural network and multiple linear regression models, Engineering Structures 33 (3) (2011) 903 – 910. doi:10.1016/j.engstruct.2010.12.011.

[28] V. Ranković, N. Grujović, D. Divac, N. Milivojević, A. Novaković, Modelling of dam behaviour based on neuro-fuzzy identification, Engineering

Structures 35 (2012) 107 – 113. doi:10.1016/j.engstruct.2011.11.011.

[29] F. Salazar, R. Morán, M. Á. Toledo, E. Oñate, Data-based models for the prediction of dam behaviour: a review and some methodological considerations, Archives of Computational Methods in Engineering 24 (1) (2017) 1–21. doi:10.1007/s11831-015-9157-9.

[30] F. Salazar, M. Á. Toledo, J. M. González, E. Oñate, Early detection of anomalies in dam performance: A methodology based on boosted regression trees, Structural Control and Health Monitoring (2017) e2012–n/aE2012 stc.2012. doi:10.1002/stc.2012.

[31] F. Salazar, M. Á. Toledo, E. Oñate, B. Suárez, Interpretation of dam deformation and leakage with boosted regression trees, Engineering Structures 119 (2016) 230 – 251. doi:10.1016/j.engstruct.2016.04.012.

[32] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Rubin, Bayesian data analysis, 3rd Edition, CRC Press, 2014.

[33] T. Hastie, R. Tibshirani, J. Friedman, J. Franklin, The elements of statistical learning: data mining, inference and prediction, The Mathematical Intelligencer 27 (2) (2005) 83–85.

[34] S. McKinley, M. Levine, Cubic spline interpolation, College of the Redwoods 45 (1) (1998) 1049–1060.

[35] S. S. Rao, A course in time series analysis, Tech. rep., Technical report, Texas A&M University (2008).