

Bayesian dynamic linear models for structural health monitoring

James-A. Goulet^{id}

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

Correspondence

James-A. Goulet, Department of Civil, Geologic and Mining Engineering, Polytechnique Montreal, Montreal, Canada.
Email: James.A.Goulet@gmail.com

Funding information

Swiss National Science Foundation; Fonds de recherche du Québec – Nature et technologies (FRQNT); National Research Council of Canada, Grant/Award Number: RGPIN-2016-06405

Summary

In several countries, infrastructure is in poor condition, and this situation is bound to remain prevalent for the years to come. A promising solution for mitigating the risks posed by ageing infrastructure is to have arrays of sensors for performing, in real time, structural health monitoring across populations of structures. This paper presents a Bayesian dynamic linear model framework for modeling the time-dependent responses of structures and external effects by breaking it into components. The specific contributions of this paper are to provide (a) a formulation for simultaneously estimating the hidden states of structural responses as well as the external effects it depends on, for example, temperature and loading, (b) a state estimation formulation that is robust toward the errors caused by numerical inaccuracies, (c) an efficient way for learning the model parameters, and (d) a formulation for handling nonuniform time steps.

KEY WORDS

Bayesian models, bridge, dynamic linear models, infrastructure, Kalman filter, structural health monitoring (SHM)

1 | INTRODUCTION

In several countries, infrastructure is in poor condition, and this situation is bound to remain prevalent for the years to come. A promising solution for mitigating the risk posed by ageing infrastructure is to have arrays of sensors deployed across cities to monitor, in real time, the condition of infrastructure. We now have the technological capacity to measure and store the data for thousands of structures. However, what is holding back structural health monitoring (SHM) is that there is currently no generic and robust way to interpret the data collected by sensors. Many authors have approached this challenge using probabilistic methods such as response surfaces,^[1] ARX,^[2] ARMA,^[3] and Kalman filter (KF)-based methods.^[4, 5] Other have approached the damage detection problem using pattern-recognition techniques^[6] and dynamic Bayesian networks.^[7] Another active subfield is dedicated to the design of sensors systems themselves.^[8–10] In their books, Yuen^[11] and Farrar and Worden^[12] present several Bayesian methods applied to SHM applications.

A key challenge remaining for enabling large-scale applications of SHM is to identify from time-series, the baseline response of structures without the effect of external actions such as temperature and traffic. To succeed at this task, a methodology must be able to operate seamlessly in conditions with frequent outliers and missing data. This paper proposes a framework for modeling the time-dependent responses of structures by breaking it into generic components, each having its own specific mathematical formulation. This new framework is a generalization of Bayesian dynamic linear models (BDLMs) for the field of SHM.

BDLMs are issued from the field of applied statistics where it is extensively used in business and finance applications.^[13, 14] The theory behind BDLM comes from the field of control where took place the development of the KF for the control system of the Apollo mission.^[15] In the field of machine learning, BDLMs are commonly referred to as state-space models.^[16] In civil engineering, several examples of applications involving the KF with structural dynamic models are published in the field of

SHM.^[4–23] A first attempt to employ a BDLM for SHM was presented by Solhjell.^[24] Solhjell introduced many new concepts of BDLM to the field of SHM, yet, several key aspects are missing for enabling the general applications on civil structures.

This paper presents a framework employing BDLM that is capable of estimating hidden state variables using time-discrete monitoring data of a structural response. Here, the term *hidden* refers to variables that are indirectly observed and *structural responses* refers to the behavior of a structure such as its frequencies, displacement, strains, and so forth. Part of the contribution of this paper is to regroup in a holistic framework, the theories found in the fields of machine learning^[16, 25] and applied statistics.^[13–24] This framework enables the general application of BDLM for the specific challenges encountered in SHM applications. The specific contributions of this paper are to provide

1. A formulation for simultaneously estimating the hidden states of structural responses as well as the external effects it depends on, for example, temperature and loading,
2. A state estimation formulation that is robust toward the errors caused by numerical inaccuracies,
3. An efficient way for learning the parameters of the model,
4. A formulation for handling nonuniform time steps.

Section 2 presents the general formulation of the BDLM as well as the generic formulation for modeling each component involved in common SHM applications. Section 3 presents the general formulation behind the KF and Kalman smoother (KS) that can be employed to estimate the hidden states of the model. This section also introduces the U-D filter that solves the numerical accuracy issues of the KF and KS. Section 4 presents two expectation maximization formulation for learning the model parameters. Finally, Section 5 presents a simulated example showcasing the BDLM.

2 | BAYESIAN DYNAMIC LINEAR MODELS

BDLMs are analogous to hidden Markov models excepted that the states are Gaussian random variables, and the state transitions are defined by linear functions. The BDLM approach presented here aims at directly modeling the responses of a structure without requiring knowledge about its structural properties. Because it requires orders of magnitudes less structure-specific knowledge than what is required for building traditional finite element models, it is trivial to build a BDLM for any type of structure.

This section presents the general formulation of BDLM, the specificities of components modeling, the formulation for exact state estimation, a graphical representation for BDLMs, and a procedure to handle nonuniform time steps.

2.1 | BDLM formulation

In a BDLM, observations \mathbf{y}_t at a time $t \in (1 : T)$ are modelled by superposing hidden states \mathbf{x}_t from several generic components as defined by the observation model

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

where the *observation matrix* \mathbf{C}_t indicates how each component from the hidden state vector \mathbf{x}_t contribute to observations \mathbf{y}_t . The dynamic evolution of the hidden states \mathbf{x}_t is described by the *transition model*

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

In Equations 1 and 2, respectively, \mathbf{v}_t denotes the Gaussian *measurement* errors with mean $\mathbf{0}$ and covariance \mathbf{R}_t , and \mathbf{w}_t denotes the Gaussian *model* errors with mean $\mathbf{0}$ and covariance \mathbf{Q}_t . Note that when the model parameters are *stationary*, the index t can be dropped for matrices A , C , R , and Q in Equations 1 and 2.

The BDLM serves three main purposes. A first purpose is to decompose a complex signal in its components by estimating the conditional probability of hidden state variables \mathbf{x}_t , given all observations up to the current time step t , $p(\mathbf{x}_t | \mathbf{y}_{1:t})$. In this paper, $p(\cdot | \cdot)$ denotes a conditional probability. A second purpose is to improve estimates of the structure behavior at a current time t , $p(\mathbf{y}_t | \mathbf{y}_{1:t})$, by combining the all the information obtained from observations up to t . A third purpose is to forecast the structure behavior at n time steps beyond the current time t , given all the information obtained up to t , $p(\mathbf{y}_{t+n} | \mathbf{y}_{1:t})$.

2.2 | BDLM components for SHM

The key aspect of BDLMs is to represent the behavior of a system by superposing the effect of each of its hidden components. The term *hidden* means that a component is not directly observed. Figure 1 presents an example of raw structural responses recorded on a structure. In this illustrative example, the raw structural response is made of the superposition of several hidden components: (a) a local level (*LL*), (b) a cyclic temperature effect (*S*), and (c) an autoregressive error term (*AR*) to account for missing physical phenomena in the model.

Subsections 2.2.1–2.2.4 present the formulation for modeling each common generic component that are useful for SHM applications. Note that each component is made of one or more hidden state variable. Here, the emphasis is put on the term *generic* because a vast array of structural behavior can be modelled by assembling combinations of the *generic components* presented in this section. The details about the derivation of this formulation is described by West and Harisson.^[14]

2.2.1 | Local level and local trend components

The local level is a basic component that is found in almost every BDLM. It represents the evolution of the baseline response of the structure without the effect of external solicitations such as traffic loading or temperature. Analogously, a local level is employed to represent the average temperature and traffic loading. The local level generic formulation is defined by

$$\mathbf{x}^{\text{LL}} = x_t^{\text{LL}}, \quad \mathbf{A}^{\text{LL}} = 1, \quad \mathbf{C}^{\text{LL}} = 1, \quad \mathbf{Q}^{\text{LL}} = (\sigma^{\text{LL}})^2$$

where the parameter σ^{LL} characterizes the model error term. This component is assumed to be *locally constant*, yet it can exhibit changes over time.

Local trend components are suited for modeling a *locally constant* rate of change in the local level. Local trends are typically employed to model a drift over time in the response of structures. The generic formulation for the local trend is

$$\mathbf{x}^{\text{LT}} = \begin{pmatrix} x_t^{\text{LL}} \\ x_t^{\text{LT}} \end{pmatrix}, \quad \mathbf{A}^{\text{LT}} = \begin{pmatrix} 1 & \Delta t \\ 0 & 1 \end{pmatrix}, \quad \mathbf{C}^{\text{LT}} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \mathbf{Q}^{\text{LT}} = (\sigma^{\text{LT}})^2 \begin{pmatrix} \frac{\Delta t^3}{3} & \frac{\Delta t^2}{2} \\ \frac{\Delta t^2}{2} & \Delta t \end{pmatrix}$$

where Δt is the time step size. In the transition matrix \mathbf{A}^{LT} , the first row adds to the local level the change cause by the local trend. The second row represents the locally constant trend. Analogously to the local level, over long periods, the local trend component can exhibit variation in the rate of change. In \mathbf{C}^{LT} , the 0 in the second row indicates that only the local level contributes to the observation y . \mathbf{Q}^{LT} describes the effect of a variation in the rate of change, on the local level and local trend. Zarchan and Musoff^[26] present the formulation for the *local acceleration* higher order model, and Mehrotra and Mahapatra^[27] present the formulation for the *local jerk* model.

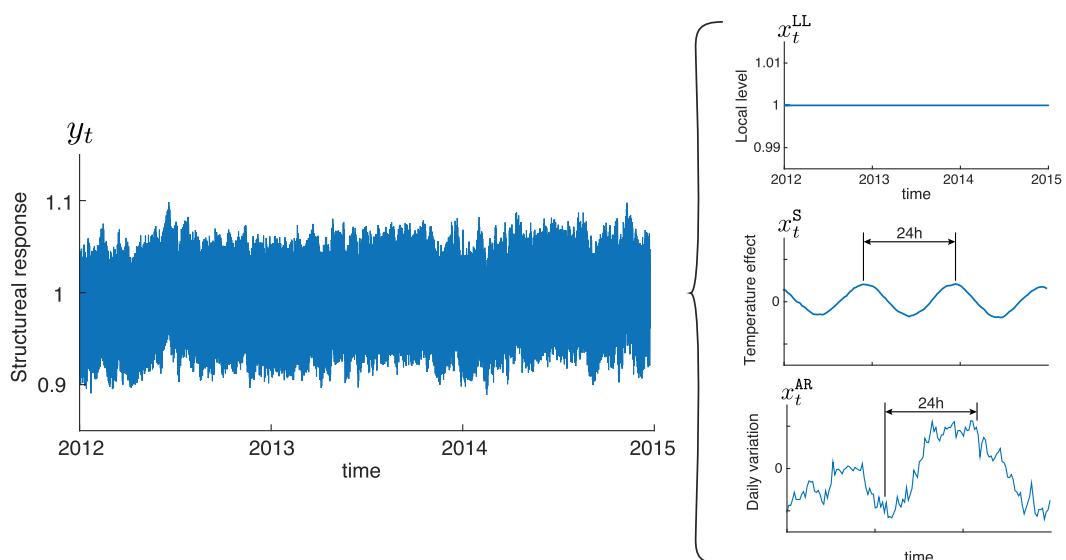


FIGURE 1 Example of raw structural response that is decomposed in three hidden components: (a) local level (*LL*), (b) cyclic temperature effect (*S*), and (c) an autoregressive error term (*AR*) to account for missing physical phenomena in the model

2.2.2 | Periodic components

Periodic components such as the daily variation ($p = 1\text{day}$) and seasonal variation ($p = 365\text{days}$) are described in their *Fourier form* by

$$\mathbf{x}^S = \begin{pmatrix} x^{S1} \\ x^{S2} \end{pmatrix}, \quad \mathbf{A}^S = \begin{pmatrix} \cos \omega & \sin \omega \\ -\sin \omega & \cos \omega \end{pmatrix}, \quad \mathbf{C}^S = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \mathbf{Q}^S = \begin{pmatrix} (\sigma^{S1})^2 & 0 \\ 0 & (\sigma^{S2})^2 \end{pmatrix}$$

where $\omega = \frac{2\pi \cdot \Delta t}{p}$, Δt is the time step length, and where x^{S1} corresponds to the amplitude of the periodic component. As indicated by the observation matrix \mathbf{C}^S , only the hidden state x^{S1} contributes to the observation y_t . For SHM applications, periodic components are suited to model daily and seasonal temperature variations that are described by two separated components. Note that this periodic component is not suited to model nonharmonic effects, for example, such as those identified by Yuen and Kuok.^[28]

2.2.3 | Autoregressive components

The autoregressive component is suited to capture the dependence of the model errors between time steps. The dependence between model errors typically arises from missing physical phenomena in the BDLM. Although the AR components may contain several terms, the most common AR component includes only one term (AR(1)) so that

$$\mathbf{x}^{\text{AR}} = x^{\text{AR}}, \quad \mathbf{A}^{\text{AR}} = \phi^{\text{AR}}, \quad \mathbf{C}^{\text{AR}} = 1, \quad \mathbf{Q}^{\text{AR}} = (\sigma^{\text{AR}})^2$$

This component exhibits two main regimes depending on the value of ϕ^{AR} .^[29] For $0 < \phi^{\text{AR}} < 1$, the AR component is a stationary process with a fixed variance given by

$$(\sigma^{\text{AR},0})^2 = \frac{(\sigma^{\text{AR}})^2}{1 - (\phi^{\text{AR}})^2}.$$

For $\phi^{\text{AR}} \geq 1$, the AR component is a nonstationary process with no fixed value mean and variance. The special case where $\phi^{\text{AR}} = 1$ corresponds to a random walk. For most SHM applications, $0 < \phi^{\text{AR}} < 1$.

2.2.4 | Regression components

Regression components are employed to describe the dependencies between the state variables associated with different observations. In the context of SHM, regression components are employed to describe the dependencies between the response of a structure and the hidden state variables associated with temperature and traffic loading observations. Unlike other components previously presented, regression components are not defined by block component matrices. Instead, a regression component adds off-diagonal terms on the global observation matrix \mathbf{C} . The relationship between the observation i and state variable j is taken into account by defining $[\mathbf{C}]_{ij} = \phi^{i|j,R}$, where $\phi^{i|j,R} \in \mathbb{R}$ is a regression coefficient.

2.3 | Graphical representation

Pearl^[30] introduced graphical models, also known as Bayesian networks, as a general representation for the joint probability of random variables. Graphical models combine the *theory of probabilities* with the *theory of graphs*. An example of trivial BDLM including only a local level is presented by a graphical model in Figure 2 where the graph in (b) is a shorthand notation for

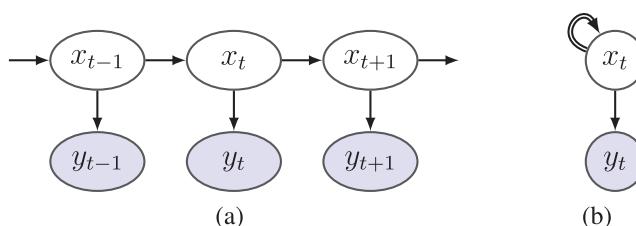


FIGURE 2 Graphical models describing (a) the expanded and (b) the compact graphical model representation of a trivial example of Bayesian dynamic linear model including only a local level component

the graph in (a). In these graphical models, circles represent random variables; links correspond to causal relations; double-line links in the graph (b) are a shorthand notation for links between time steps in the graph (a). Color-filled nodes correspond to *observed variables*; white-filled nodes correspond to *unobserved/hidden variables*.

2.4 | Nonuniform time steps

One challenge that arises while working with real data is that the time steps when data is recorded can be nonuniformly spaced. In order to accommodate nonuniform time steps, a reference time step Δt^{ref} must be defined. All parameter values in the parameter set \mathcal{P} will be estimated for this reference time step. Parameter values for time steps that have a length Δt different than the reference value will have to be adapted.

For parameters associated with additive modeling error terms, that is, model error standard deviations in \mathbf{Q} , it is proposed to scale σ proportionally to the ratio between the current time step and the reference time step so that

$$\sigma^{\Delta t} = \sigma^{\Delta t, \text{ref}} \frac{\Delta t}{\Delta t^{\text{ref}}}.$$

For the model transition matrix \mathbf{A} , the components associated with \mathbf{A}^S and \mathbf{A}^{AR} also need to be adapted. For \mathbf{A}^S , the time step Δt must be modified in the angular frequency term ω . \mathbf{A}^{AR} contains the autoregressive coefficients that are recursively multiplied with the hidden state at each time step. To account for time step changes, autoregressive coefficients ϕ^{AR} are elevated to the power of the ratio between the current time step and the reference time step

$$\phi^{\text{AR}, \Delta t} = (\phi^{\text{AR}, \Delta t, \text{ref}})^{\frac{\Delta t}{\Delta t^{\text{ref}}}}.$$

When a local trend component is employed, Δt should be adapted in \mathbf{A}^{LT} . Note that the parameters in the observation matrix \mathbf{C} and the measurement noise matrix \mathbf{R} remain unchanged when the length of a time step is modified.

3 | BDLM STATE ESTIMATION: KALMAN AND U-D FILTERS/SMOOTHERS

For a time series where $t = 1 : T$, the most common algorithms for estimating the hidden state variables are the KF and KS. The KF is suited for the online (i.e., in real time) estimation of $p(\mathbf{x}_t | \mathbf{y}_{1:t})$, the hidden states at a time t given available observations $\mathbf{y}_{1:t}$.

The KF algorithm is commonly divided into the *prediction* (Equation 3) and *measurement* (Equation 4) steps.

$$\begin{aligned} p(\mathbf{x}_t | \mathbf{y}_{1:t-1}) &= \int \mathcal{N}(\mathbf{x}_t | \mathbf{A}_t \mathbf{x}_{t-1}, \mathbf{Q}_t) \mathcal{N}(\mathbf{x}_{t-1} | \boldsymbol{\mu}_{t-1}, \Sigma_{t-1}) d\mathbf{x}_{t-1} \\ &= \mathcal{N}(\mathbf{x}_t | \boldsymbol{\mu}_{t|t-1}, \Sigma_{t|t-1}) \\ \boldsymbol{\mu}_{t|t-1} &\triangleq \mathbf{A}_t \boldsymbol{\mu}_{t-1} \\ \Sigma_{t|t-1} &\triangleq \mathbf{A}_t \Sigma_{t-1} \mathbf{A}_t^\top + \mathbf{Q}_t \end{aligned} \quad (3)$$

$$\begin{aligned} p(\mathbf{x}_t | \mathbf{y}_{1:t}) &= \mathcal{N}(\mathbf{x}_t | \boldsymbol{\mu}_t, \Sigma_t) \\ \boldsymbol{\mu}_t &= \boldsymbol{\mu}_{t|t-1} + \mathbf{K}_t \mathbf{r}_t && \text{Posterior expected value} \\ \Sigma_t &= (\mathbf{I} - \mathbf{K}_t \mathbf{C}_t) \Sigma_{t|t-1} && \text{Posterior covariance} \\ \mathbf{r}_t &\triangleq \mathbf{y}_t - \hat{\mathbf{y}}_t && \text{Innovation vector} \\ \hat{\mathbf{y}}_t &\triangleq \mathbb{E}[\mathbf{y}_t | \mathbf{y}_{1:t-1}] = \mathbf{C}_t \boldsymbol{\mu}_{t|t-1} && \text{Prediction observations vector} \\ \mathbf{K}_t &\triangleq \Sigma_{t|t-1} \mathbf{C}_t^\top \mathbf{S}_t^{-1} && \text{Kalman gain matrix} \\ \mathbf{S}_t &\triangleq \mathbf{C}_t \Sigma_{t|t-1} \mathbf{C}_t^\top + \mathbf{R}_t && \text{Innovation covariance matrix} \end{aligned} \quad (4)$$

The *Kalman gain* \mathbf{K}_t represents the ratio between the prior covariance defined by the dynamic model and the measurement error covariance. If the variance of the prior knowledge is small (large) and the variance of the measurement errors large (small), the Kalman gain tends to one (zero). In the case of the posterior expected value, the Kalman gain is employed to weight the information coming from the prior $\boldsymbol{\mu}_{t|t-1}$ and the information coming from the observations, through the innovation vector \mathbf{r}_t .

The KS is suited for offline estimation of the state at any time t given the entire time series $\mathbf{y}_{1:T}$. The KS is initialized with the last step of the KF (μ_{TT}, Σ_{TT}), and it propagates the information obtained for time steps $t : T$, on the state estimates for all previous time steps $1 : t - 1$. The KS formulation follows

$$\begin{aligned} p(\mathbf{x}_t | \mathbf{y}_{1:T}) &= \mathcal{N}(\mathbf{x}_t | \boldsymbol{\mu}_{t|T}, \Sigma_{t|T}) \\ \boldsymbol{\mu}_{t|T} &= \boldsymbol{\mu}_{t|t} + \mathbf{J}_t (\boldsymbol{\mu}_{t+1|T} - \boldsymbol{\mu}_{t+1|t}) && \text{Posterior expected value} \\ \Sigma_{t|T} &= \Sigma_{t|t} + \mathbf{J}_t (\Sigma_{t+1|T} - \Sigma_{t+1|t}) \mathbf{J}_t^\top && \text{Posterior covariance} \\ \mathbf{J}_t &\triangleq \Sigma_{t|t} \mathbf{A}_{t+1}^\top \Sigma_{t+1|t}^{-1} && \text{Backward Kalman gain matrix} \end{aligned} \quad (5)$$

Although the KF/KS formulations presented in Equations 3 to 5 are the most widespread, it is also known to suffer from numerical instability issues,^[15] especially when

1. The covariance is rapidly reduced in the measurement step, such as when extremely accurate measurements are used or when data is used after a long period where data was missing.
2. There is a large difference between the variance of several state variables contributing to the same observation.

For SHM applications, these cases are common and cause numerical instability issues. The U-D filter proposed by Bierman and Thornton^[31] solves this issue by factorizing the covariance matrix using the Cholesky decomposition so that

$$\Sigma_t = \mathbf{U}_t \mathbf{D}_t \mathbf{U}_t^\top$$

Besides being numerically more stable, the U-D filter computes the same quantities, that is, $p(\mathbf{x}_t | \mathbf{y}_{1:t})$ and has the same physical interpretation as the KF. Moreover, it also has a comparable computational efficiency.^[15] Because the formulation of the U-D filter is more involved than the one from the KF, the reader is invited to refer to specialized literature such as Gibbs^[15] and Simon^[32] for the complete formulation.

4 | BDLM PARAMETER ESTIMATION: EM ALGORITHM

A key aspect of BDLM is the estimation of model parameters for each component. One way of estimating parameters is to use an Expectation Maximisation (EM) algorithm.

4.1 | E-step: Log-likelihood estimation using the KS

The E in E-step stands for expectation. The estimation of model parameters requires the definition of a *performance metric* for the model. The conventional performance metric is the log-likelihood of observations. The log-likelihood of an observation is the logarithm of the prior probability of this observation at a time t , given our state estimate at time $t - 1$. Based on the hypothesis that all observations are independent, the log-likelihood is defined as

$$\ln p(\mathbf{y}_{1:T}) = \sum_{t=1}^T \ln (\mathcal{N}(\mathbf{y}_t | \mathbf{C}_t \boldsymbol{\mu}_{t|t-1}, \mathbf{S}_t)) \quad (6)$$

where for the purpose of numerical accuracy, the product of the probabilities $\mathcal{N}(\mathbf{y}_t | \mathbf{C}_t \boldsymbol{\mu}_{t|t-1}, \mathbf{S}_t)$ is transformed as the sum of the log of probabilities. In the E-step, the log-likelihood is computed based on the state estimates $p(\mathbf{x}_t | \mathbf{y}_{1:t-1})$ obtained using a filtering algorithm.

4.2 | M-step: Gradient ascent

The M in M-step stands for maximization. The goal of the maximization step is to identify parameter values that maximize the log-likelihood estimated during a training period.

In this paper, two approaches are presented for the M-step, each having their strength and weaknesses. The first maximization scheme evaluates the derivative of the log-likelihood (see Equation (6)) with respect to each matrix ($\mathbf{A}^{\text{old}}, \mathbf{C}^{\text{old}}, \mathbf{Q}^{\text{old}}, \mathbf{R}^{\text{old}}$) and sets it to zero to identify the new matrices ($\mathbf{A}^{\text{new}}, \mathbf{C}^{\text{new}}, \mathbf{Q}^{\text{new}}, \mathbf{R}^{\text{new}}$) that are maximizing the log-likelihood. Ghahramani and Hinton^[25] have derived the analytical formulation resulting from the maximization operation so that

$$\begin{aligned}
 \mathbf{C}^{\text{new}} &= \left(\sum_{t=1}^T \mathbf{y}_t \boldsymbol{\mu}_t^\top \right) \left(\sum_{t=1}^T \boldsymbol{\Sigma}_t \right)^{-1} \\
 \mathbf{R}^{\text{new}} &= \frac{1}{T} \sum_{t=1}^T (\mathbf{y}_t \mathbf{y}_t^\top - \mathbf{C}^{\text{new}} \boldsymbol{\mu}_t \boldsymbol{\mu}_t^\top) \\
 \mathbf{A}^{\text{new}} &= \left(\sum_{t=2}^T \boldsymbol{\Sigma}_{t,t-1} \right) \left(\sum_{t=2}^T \boldsymbol{\Sigma}_{t-1} \right)^{-1} \\
 \mathbf{Q}^{\text{new}} &= \frac{1}{T-1} \left(\sum_{t=2}^T \boldsymbol{\Sigma}_t - \mathbf{A}^{\text{new}} \sum_{t=2}^T \boldsymbol{\Sigma}_{t,t-1}^\top \right) \\
 \boldsymbol{\Sigma}_{t,t-1} &= (\mathbf{I} - \mathbf{K}_t \mathbf{C}_t) (\mathbf{A}_t \boldsymbol{\Sigma}_{t-1}).
 \end{aligned} \tag{7}$$

In addition to the model matrices, the initial hidden state can be updated following

$$\begin{aligned}
 \boldsymbol{\mu}_1^{\text{new}} &= \boldsymbol{\mu}_1 \\
 \boldsymbol{\Sigma}_1^{\text{new}} &= \boldsymbol{\Sigma}_1 - \boldsymbol{\mu}_1 \boldsymbol{\mu}_1^\top.
 \end{aligned} \tag{8}$$

Note that in Equations 7 and 8, $\boldsymbol{\mu}_t$ and $\boldsymbol{\Sigma}_t$ are estimated using a smoothing algorithm. The maximization operations presented in Equation 7 modifies simultaneously all terms in updated matrices (\mathbf{A}^{new} , \mathbf{C}^{new} , \mathbf{Q}^{new} , \mathbf{R}^{new}). It is thus necessary to replace specific terms that are not unknown parameters by the initial values in (\mathbf{A}^{old} , \mathbf{C}^{old} , \mathbf{Q}^{old} , \mathbf{R}^{old}). Optimizing simultaneously all the model matrices in order to maximize the log-likelihood has the upside to be fast and computationally efficient. However, this approach is also known for getting trapped in local maxima.^[33] Because matrices \mathbf{A} , \mathbf{C} , \mathbf{Q} , \mathbf{R} are typically sparse, modifying simultaneously all the terms in these matrices can lead to undesirable local maxima.

A second approach that can overcome this limitation is to perform the maximization with a parameter-wise approach instead of matrix-wise one. This maximization scheme employs a Newton-Raphson approach^[34] where the first and second derivative of the log-likelihood (see Equation (6)) with respect to each parameter θ_i enables to find new optimized parameter values so that

$$\theta_i^{\text{new}} = \theta_i^{\text{old}} - \frac{\frac{\partial \ln p(\mathbf{y}_{1:T})}{\partial \theta_i}}{\frac{\partial^2 \ln p(\mathbf{y}_{1:T})}{\partial^2 \theta_i}}.$$

Because this approach only optimizes the parameters of interest instead of entire matrices, it provides an additional tool in case the matrix-wise maximization is trapped in a local maximum.

The general procedure is to repeat the E and M steps recursively until the change in log-likelihood between iterations is below an admissible threshold. The main limitation of the EM algorithm is that it is only guaranteed to converge to a local maximum. This is an issue because the log-likelihood function is usually nonconvex, which leads to several local maxima. One mitigation strategy is to use random initial parameter values to search for the region containing the global maximum.

5 | ILLUSTRATIVE EXAMPLE

An illustrative example showcases (a) the construction of a BDLM, (b) model parameter estimation, (c) the separation of raw observations in generic components, and (d) the effect of missing data and outliers.

5.1 | Simulated data

For the purpose of this example, simulated data is generated for representing a typical SHM application. The behavior of structures is often affected by temperature changes as well as traffic loading. Temperature variation is affected by a daily and a seasonal cycle. Traffic variation is affected by the daily traffic pattern. The structure's behavior that is simulated here is the first resonant frequency. All datasets are modelled for discrete time steps at intervals of 30 min (i.e., $\Delta t = \frac{1}{48}$ days/observations) for a total duration of 2 years so that $t = 1, 2, \dots, T$; $T = 2 \times 365 \times 48$. Figure 3 presents the simulated observations generated for this example.

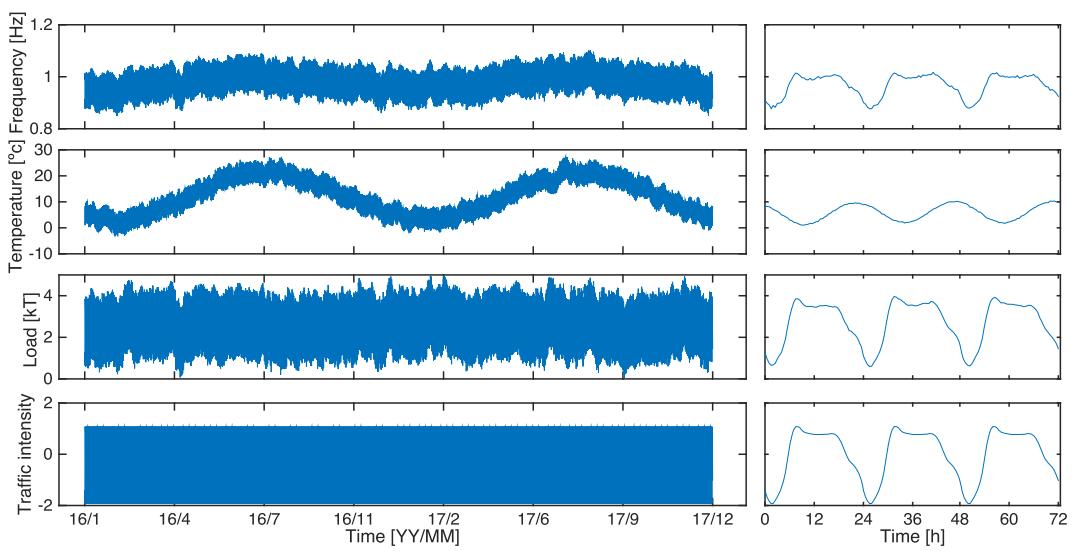


FIGURE 3 Simulated data for the illustrative example. The left graphs represent the entire 2 years dataset and the right graphs represent only the first 3 days

The detailed procedure employed to generate simulated observations is detailed below. Note that the superscript nomenclature follows $(^T)$ temperature, $(^P)$ traffic pattern, $(^L)$ traffic load, and $(^B)$ frequencies. Also, following the nomenclature presented in Section 2.2: $(^{LL})$ local level, $(^S)$ cyclic component, $(^{AR})$ autoregressive component, and $(^R)$ regression component.

5.1.1 | Temperature – y_t^T

The simulated temperature observations y_t^T are obtained by superposing the average temperature, daily ($p = 1$ day) and seasonal ($p = 365$ day) sinusoidal cycles, an autoregressive (AR(1)) process, and measurement errors so that

$$y_t^T = x_t^{T, LL} + x_t^{T, S_1} + x_t^{T, S_2} + x_t^{T, AR} + v_t^T \quad (9)$$

where each component is defined following

$$\begin{aligned} x_t^{T, LL} &= 12^\circ C && \text{(average temperature)} \\ x_t^{T, S_1} &= 4 \sin\left(\frac{2\pi}{1}\left(\frac{t}{48} + \frac{8}{24}\right)\right) && \text{(daily cycle)} \\ x_t^{T, S_2} &= 9 \sin\left(\frac{2\pi}{365}\left(\frac{t}{48} + \frac{8}{12} \cdot 365\right)\right) && \text{(seasonal cycle)} \\ x_t^{T, AR} &= \underbrace{0.995}_{\phi^{T, AR}} \cdot x_{t-1}^{T, AR} + w_t^{T, AR}, \quad w_t^{T, AR} \sim \mathcal{N}(0, \underbrace{0.1}_{\sigma^{T, AR}})^2 && \text{(AR(1) process)} \\ v_t^T &\sim \mathcal{N}(0, \underbrace{0.1}_{\sigma^T})^2 && \text{(measurement error)} \end{aligned} \quad (10)$$

5.1.2 | Traffic pattern – y_t^P

The traffic pattern describes the normalized traffic intensity. It is employed as a regression variable for defining the traffic load. The traffic pattern observations y_t^P are considered to be exacts and are defined following

$$y_t^P = x_t^P = h\left(\left(\frac{t}{48} - \left\lfloor \frac{t}{48} \right\rfloor\right) \cdot 24\right) \quad (11)$$

where $\lfloor \cdot \rfloor$ denotes the floor operator (i.e., round to the lowest integer) and $h(\cdot)$ is the normalized daily traffic intensity presented in Figure 4.

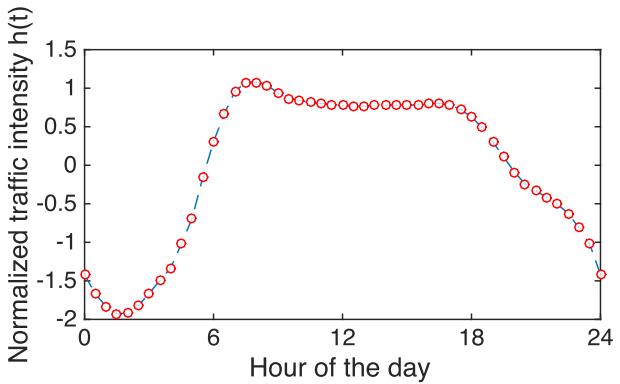


FIGURE 4 Synthetic data for daily traffic intensity

5.1.3 | Load – y_t^L

The load applied on the bridge is defined by superposing the average traffic load, the traffic intensity, which is a constant, time the traffic pattern, an autoregressive process, and observation errors, so that

$$y_t^L = x_t^{L, LL} + x_t^{L|P, R} + x_t^{L, AR} + v_t^L \quad (12)$$

where each component is defined following

$$\begin{aligned} x_t^{L, LL} &= 3 kT && \text{(average traffic load)} \\ x_t^{L|P, R} &= \underbrace{1.1}_{\phi^{L|P, R}} \cdot x_t^P && \text{(traffic intensity regression)} \\ x_t^{L, AR} &= \underbrace{0.995}_{\phi^{L, AR}} \cdot x_{t-1}^{L, AR} + w_t^{L, AR}, \quad w_t^{L, AR} \sim \mathcal{N}(0, \underbrace{0.025^2}_{\sigma^{L, AR}}) && \text{(AR(1) process)} \\ v_t^L &\sim \mathcal{N}(0, \underbrace{0.01^2}_{\sigma^L}) && \text{(measurement error)} \end{aligned} \quad (13)$$

5.1.4 | Frequencies – y_t^B

Frequency observations are simulated by superposing the baseline level of 1Hz, the effect of traffic loading on the frequency, the effect of temperature on frequency, an autoregressive component, and observation errors. Frequency observations are defined by

$$y_t^B = x_t^{B, LL} + x_t^{B|L, R} + x_t^{B|T, R} + x_t^{B, AR} + v_t^B \quad (14)$$

where each component is defined following

$$\begin{aligned} x_t^{B, LL} &= 1 \text{ Hz} && \text{(average frequency)} \\ x_t^{B|L, R} &= \underbrace{0.05}_{\phi^{B|L, R}} (x_t^{L|P, R} + x_t^{L, AR}) && \text{(traffic load regression)} \\ x_t^{B|T, R} &= \underbrace{0.0029}_{\phi^{B|T, R}} (x_t^{T_1, S_1} + x_t^{T_2, S_1} + x_t^{T, AR}) && \text{(temperature regression)} \\ x_t^{B, AR} &= \underbrace{0.995}_{\phi^{B, AR}} \cdot x_{t-1}^{B, AR} + w_t^{B, AR}, \quad w_t^{B, AR} \sim \mathcal{N}(0, \underbrace{0.0005^2}_{\sigma^{B, AR}}) && \text{(AR(1) process)} \\ v_t^B &\sim \mathcal{N}(0, \underbrace{0.0025^2}_{\sigma^B}) && \text{(measurement error)} \end{aligned} \quad (15)$$

5.2 | BDLM construction

The BMLM takes the form expressed in Equations 1 and 2 where the global matrices defining the model are

$$\begin{aligned}\mathbf{y}_t &= [y_t^B, y_t^T, y_t^L, y_t^P]^\top \\ \mathbf{x}_t &= [\mathbf{x}_t^B, \mathbf{x}_t^T, \mathbf{x}_t^L, x_t^P]^\top \\ \mathbf{A}_t &= \text{blockdiag}(\mathbf{A}_t^B, \mathbf{A}_t^T, \mathbf{A}_t^L, A_t^P) \\ \mathbf{C}_t &= \text{blockdiag}(\mathbf{C}_t^B, \mathbf{C}_t^T, \mathbf{C}_t^L, C_t^P) \\ \mathbf{R}_t &= \text{blockdiag}(R_t^B, R_t^T, R_t^L, R_t^P) \\ \mathbf{Q}_t &= \text{blockdiag}(\mathbf{Q}_t^B, \mathbf{Q}_t^T, \mathbf{Q}_t^L, Q_t^P)\end{aligned}$$

where each component of these model matrices is defined below. Complete matrices are presented in Appendix A. Model parameters to be estimated are regrouped in the set

$$\mathcal{P} = \left\{ \underbrace{\phi^{B|L,R}, \phi^{B|T,R}, \phi^{B|P,R}, \phi^{L|P,R}}_{\text{regression coefficients}}, \underbrace{\phi^{B,AR}, \phi^{T,AR}, \phi^{L,AR}}_{\text{autocorr. coefficients}}, \underbrace{\sigma^{B,AR}, \sigma^{L,AR}, \sigma^{T,AR}}_{\sqrt{\text{autocorr. variance}}} \right\}.$$

Note that all local level standard deviations, $\sigma^{B,LL} = \sigma^{L,LL} = \sigma^{T,LL} \equiv 0$ because the structure behavior, the temperature, and the load are stationary.

5.2.1 | Temperature – y_t^T

There are four components involved in the temperature model: (a) a local level, (b–c) two periodic signals with a period of one day ($\omega^{T1} = \frac{2\pi}{48}$) and 365 days ($\omega^{T2} = \frac{2\pi}{365 \cdot 48}$), (d) an autoregressive process. The vector of hidden state variables for the temperature is

$$\mathbf{x}_t^T = \left[\underbrace{x_t^{T,LL}}_{\text{local level}}, \underbrace{x_t^{T1,S1}, x_t^{T1,S2}}_{\text{cycle, } p=1 \text{ day}}, \underbrace{x_t^{T2,S1}, x_t^{T2,S2}}_{\text{cycle, } p=365 \text{ day}}, \underbrace{x_t^{T,AR}}_{\text{AR process}} \right],$$

where the ordering of each component remains the same for other matrices \mathbf{C}_t^T , \mathbf{A}_t^T , and \mathbf{Q}_t^T . The observation matrix is defined following

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

where the zeros indicate that components $x_t^{T1,S2}$ and $x_t^{T2,S2}$ do not contribute to the temperature observations. The measurement noise variance is $R_t^T = (\sigma_T)^2$. The transition matrix is

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

The model error covariance is

$$\mathbf{Q}_t^T = \text{block diag} (0, 0, 0, (\sigma^{T,AR})^2).$$

5.2.2 | Traffic pattern – y_t^P

The traffic pattern observations y_t^P are exact so that the transition model A_t^P should bear no weight during the estimation. This is achieved by using an observation variance tending to zero and a model error variance tending to infinity. With these values, the Kalman gain tends to one, so that all the weight is put on the observation rather than on the model. This strategy is employed

so that the other components in the BDLM can directly be modelled as a function of the traffic pattern observations. Matrices defining the traffic pattern component are

$$\mathbf{x}_t^P = x_t^P = y_t^P, \quad A_t^P = 1, \quad R_t^P \rightarrow 0, \quad Q_t^T \rightarrow \infty.$$

5.2.3 | Traffic load – y_t^L

There are two independent components involved in the traffic load: (a) a local level and (b) an autoregressive process. There is also one dependent component, a regression component linking the traffic pattern to the traffic load. The vector of state variables for the traffic load is

$$\mathbf{x}_t^L = \begin{bmatrix} x_t^{L, LL}, & x_t^{L, AR} \\ \text{local level} & \text{AR process} \end{bmatrix}$$

where the ordering of each component remains the same for other matrices \mathbf{C}_t^L , \mathbf{A}_t^L , and \mathbf{Q}_t^L . The observation matrix is $\mathbf{C}_t^L = [1, 1]$, the measurement noise variance is $R_t^L = (\sigma_L)^2$, the transition matrix is $\mathbf{A}_t^L = \text{blockdiag}(1, \phi^{L, AR})$, and the model error covariance is $\mathbf{Q}_t^L = \text{blockdiag}(0, (\sigma^{L, AR})^2)$. The dependent component is taken into account by introducing a regression coefficient in the global observation matrix

$$[\mathbf{C}_t]_{3,11} = \phi^{L|P, R}.$$

This includes the effect of the traffic pattern in the traffic load observation. Note that the position (3, 11) in the global observation matrix is defined by the ordering chosen for state variables.

5.2.4 | Frequency – y_t^B

There are two independent components involved in the frequency: (a) a local level and (b) an autoregressive process. There are also five dependent components: (c–h) the five regression components respectively describe the dependency between the frequency and temperature or traffic components. The vector of state variables for the frequency is

$$\mathbf{x}_t^B = \begin{bmatrix} x_t^{B, LL}, & x_t^{B, AR} \\ \text{local level} & \text{AR process} \end{bmatrix}$$

where the ordering of each component remains the same for other matrices \mathbf{C}_t^B , \mathbf{A}_t^B , and \mathbf{Q}_t^B . The observation matrix is $\mathbf{C}_t^B = [1, 1]$, the transition matrix is $\mathbf{A}_t^B = \text{blockdiag}(1, \phi^{B, AR})$ model error covariance is $\mathbf{Q}_t^T = \text{blockdiag}(0, (\sigma^{B, AR})^2)$. The dependent components are taken into account by introducing regression coefficients in the global observation matrix where

$$[\mathbf{C}_t]_{1,11} = \phi^{B|P, R}, \quad [\mathbf{C}_t]_{1,10} = \phi^{B|L, R}, \quad [\mathbf{C}_t]_{1,4} = [\mathbf{C}_t]_{1,6} = [\mathbf{C}_t]_{1,8} = \phi^{B|T, R}$$

so that the effect of temperature and traffic pattern is included in the frequency observation. The measurement noise variance is assumed to be known $R_t^B = (\sigma_B)^2$. In practice, methods such as the one proposed by Yuen and Kuok^[35] can be employed for estimating this parameters.

5.2.5 | Graphical model representation

The BDLM defined in Section 5.2 can be represented by a graphical model as described in Section 2.3. Figure 5 presents one time-slice from the dynamic Bayesian network (i.e., graphical model) describing the BDLM. This figure regroups all components presented above in Section 5.2.

5.3 | Computational efficiency

It takes approximately 0.0008 s to run one time-slice of the U-D filter on a laptop computer, and it takes approximately 15 s to process 1 year worth of data (for a sampling period of 30 min). For the same configuration, it takes approximately 0.0005 s per

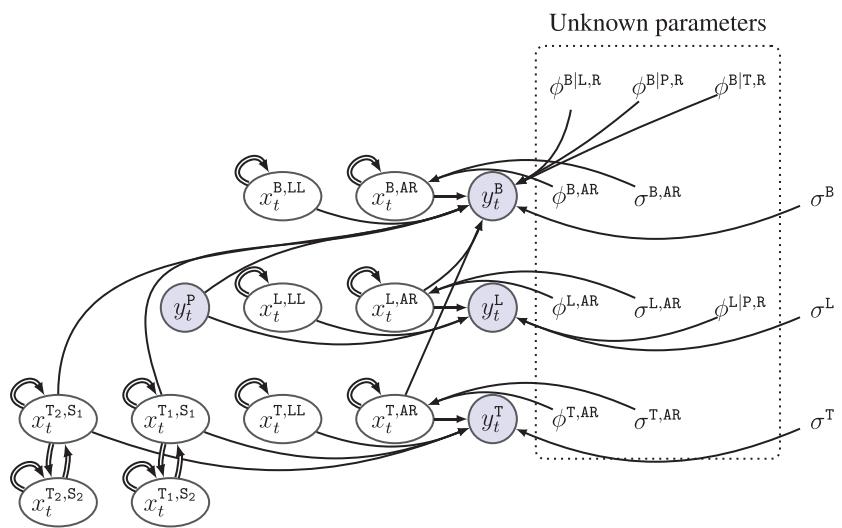


FIGURE 5 Graphical model (i.e., dynamic Bayesian network) representing the causal dependencies between each component of the model. Circles represent random variables; links correspond to causal relations; double-line links are a shorthand notation for links between time steps. Nodes without a border represent deterministic parameters; color-filled nodes correspond to *observed variables*; white-filled nodes correspond to *unobserved/hidden variables*

time slice with the standard KF presented in Section 3. Only the U-D filter formulation is employed in this paper because, even if the KF is faster, it leads to several numerical inaccuracies invalidating the final results. However, note that both the U-D and KFs are estimating the exact same quantities and have the same physical interpretation. With either the U-D or the KF, a single off-the-shelf computer would be able to process in real time, the data from thousands of sensors that would be positioned across a population of structures.

5.4 | Model parameter estimation

A training period of 6 months is employed to learn model parameters \mathcal{P} . This duration is chosen because it displays the full amplitude of the seasonal temperature (January–June). Note that without the effect of seasonal temperature, a shorter training period could be selected. The initial state \mathbf{x}_0 is described by a broad (i.e., noninformative) prior where each component has a mean of $[\mu_0]_i = 0, \forall i$ and a covariance $[\Sigma_0]_{ii} = 10^6, \forall i, [\Sigma_0]_{ij} = 0, \forall i \neq j$.

In order to establish a reference, the matrix-wise EM algorithm is first employed while starting from the true parameter values. In this configuration, the EM algorithm converges to parameter values that are negligibly close to the true values. The log-likelihood of this set of reference parameters is -3434 . Note that although the initial state \mathbf{x}_0 can be identified at the same time as other unknown parameters, it has been chosen to keep them as a broad prior in order to illustrate how information about the hidden states is gained during the training period.

In a second step, a realistic scenario is tested where 10 sets of initial parameter values are generated from a Gaussian distribution centered on true parameter values and with a coefficient of variation of 1. The EM algorithm is employed iteratively (≈ 7 seconds/evaluation) either until reaching 500 evaluations, or when the difference in the likelihood between two consecutive time steps is below 10^{-5} . Note that during the initial perturbation and optimization steps, parameter values are constrained to their feasible domain: \mathbb{R}^+ for variances (σ^2), $(0, 1)$ for the AR coefficients ϕ^{AR} , and \mathbb{R} for the regression coefficients ϕ^{ilj} .

Figure 6a presents the evolution of the log-likelihood for the starting point having led to the highest log-likelihood (-3379) and the estimated parameter values

$$\mathcal{P}^* = \{0.0501, 0.0030, 0.0539, 1.0780, 0.9964, 0.9934, 0.9935, 0.0005, 0.0100, 0.0994\}.$$

Figure 6b presents the evolution of the relative error in the parameter values. The absolute relative difference between the true parameter values and \mathcal{P}^* is at most 2.8% with a mean of 0.25%. In order to be representative of real situations where the true values remain unknown, all the following results are calculated using the parameter values \mathcal{P}^* instead of the true values.

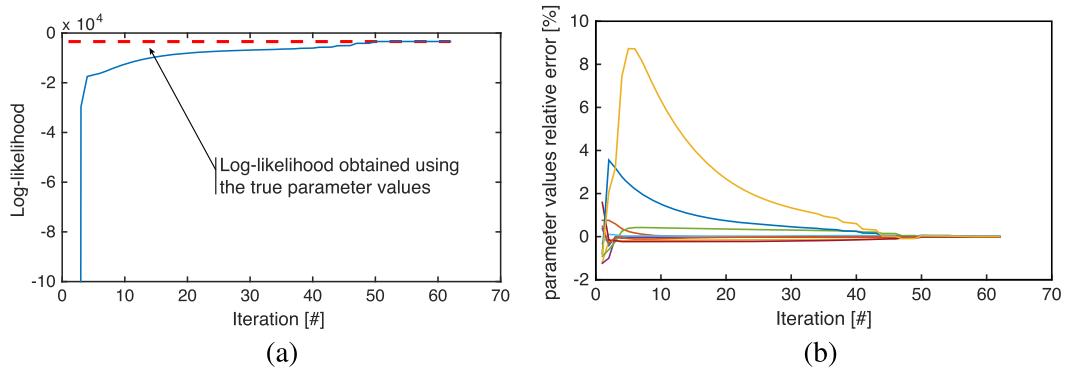


FIGURE 6 (a) Evolution of the log-likelihood with the number of Expectation Maximisation iterations for the best starting point that has led to the model parameter \mathcal{P}^* . (b) Evolution of the relative error between true and parameter values. The right most values in (b) represent the relative discrepancy between the true parameter values and \mathcal{P}^*

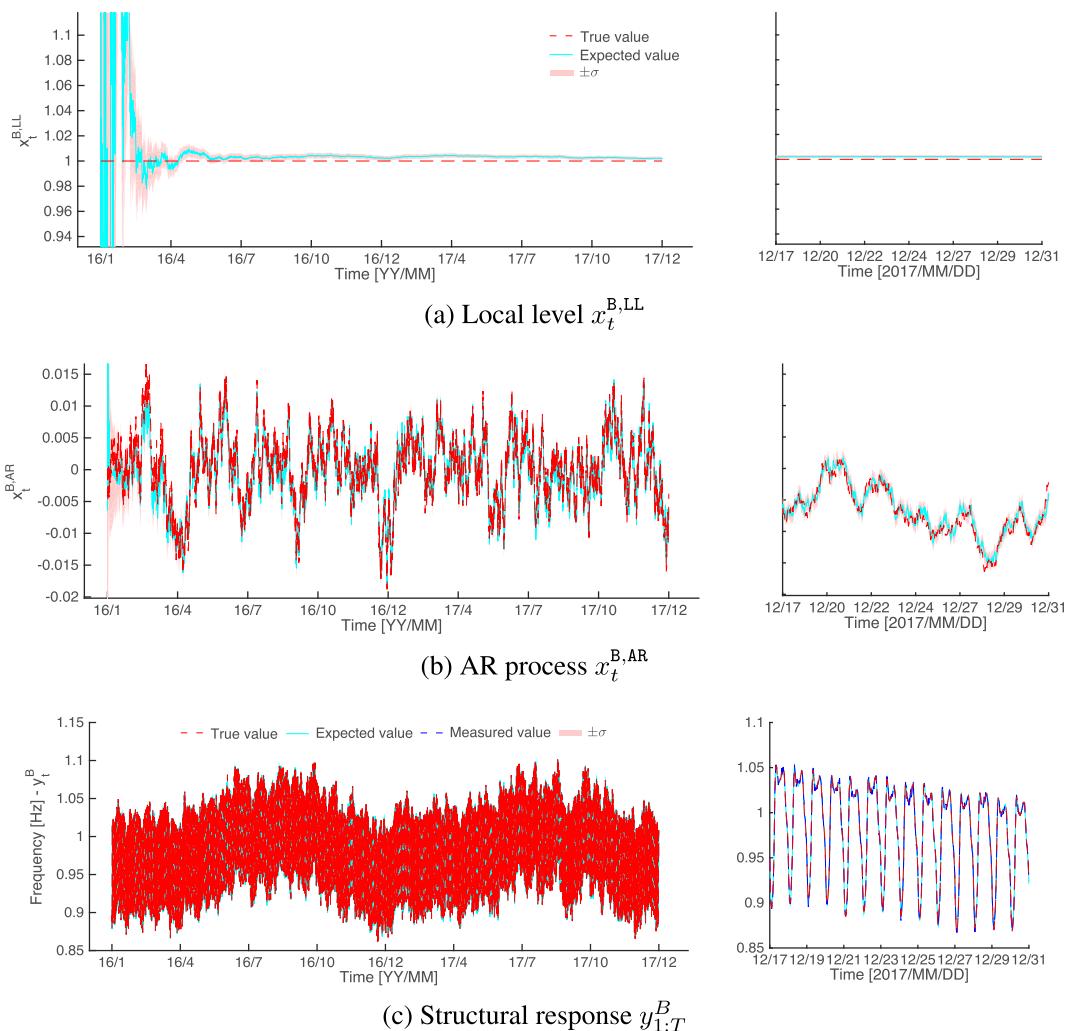


FIGURE 7 Illustration of the model component separation and filtered predictions for the structural response. All left graphs present the entire dataset (2 years) and the right graphs present the last 2 weeks

5.4.1 | Model components separation

The BDLM is employed to separate the observed signal into its hidden components. The filtered signals $\mathbf{y}_{1:T}$ along with selected hidden components are presented in Figures 7–9 where all left graphs present the entire dataset (2 years) and the right graphs

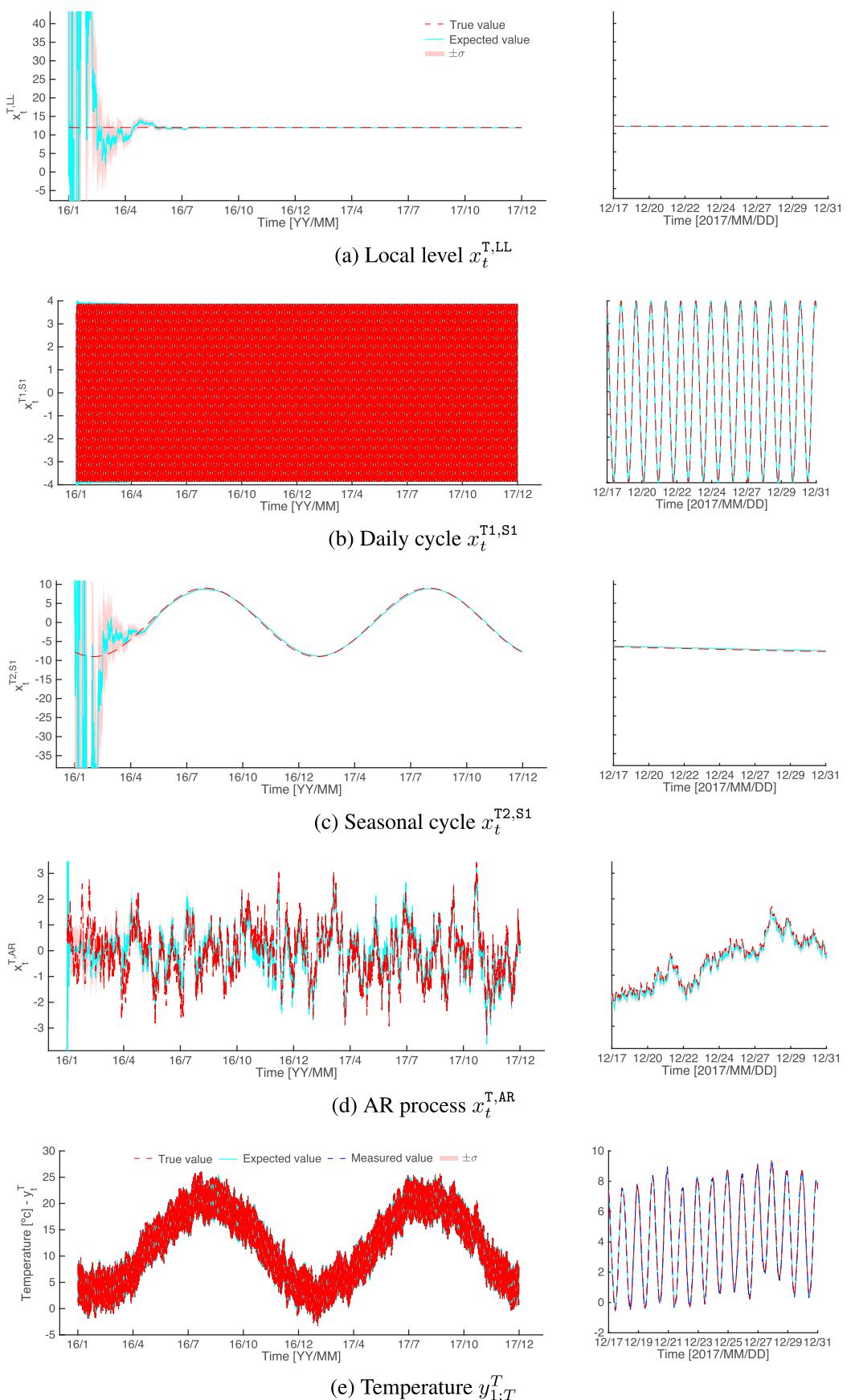


FIGURE 8 Illustration of the model component separation and filtered predictions for the temperature. All left graphs present the entire dataset (2 years) and the right graphs present the last 2 weeks

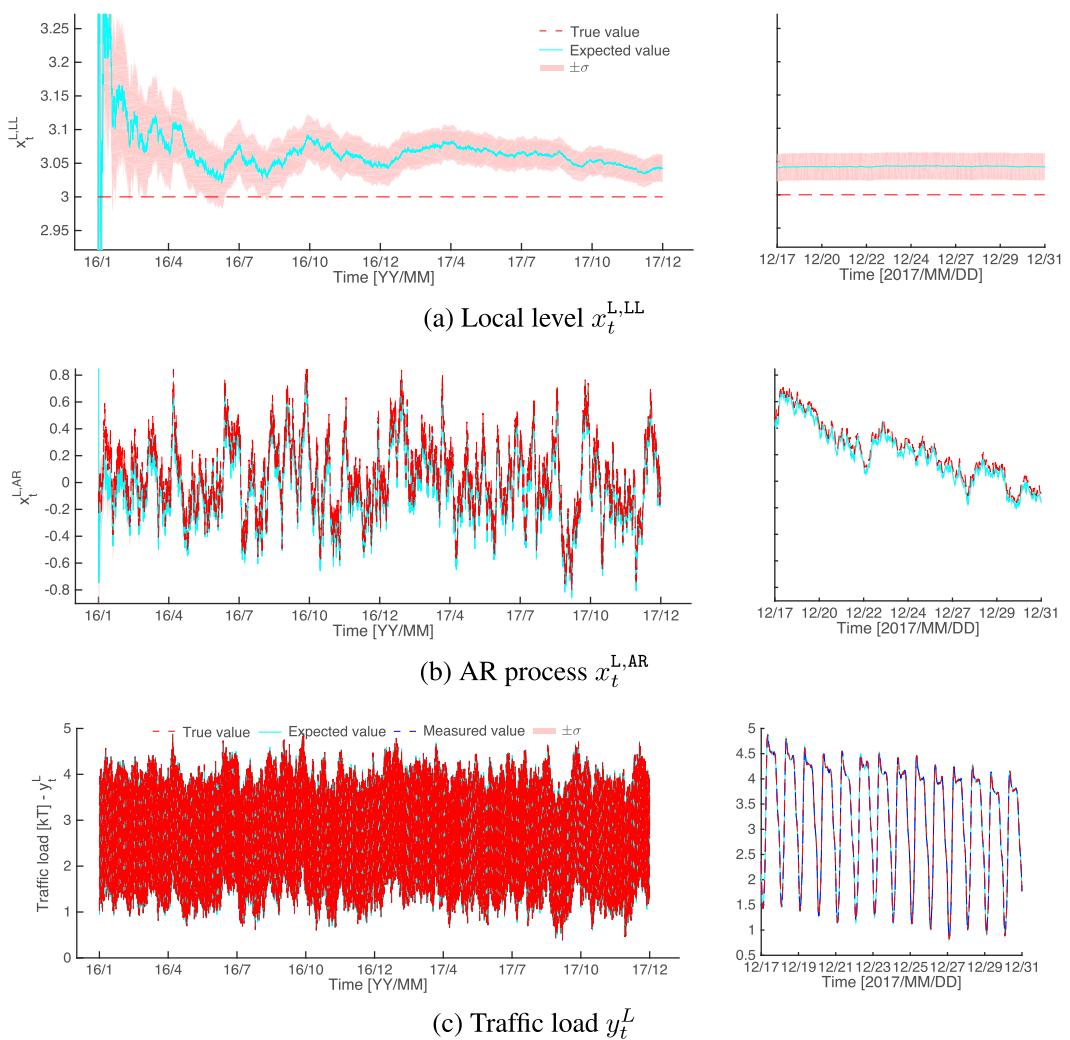


FIGURE 9 Illustration of the model component separation and filtered predictions for the traffic load. All left graphs present the entire dataset (2 years) and the right graphs present the last 2 weeks

present the last 2 weeks. Note that $x_t^{T1,S2}$ and $x_t^{T2,S2}$ are not presented because they do not directly contribute to the filtered signal $\mathbf{y}_{1:T}^T$. The load pattern regressor $\mathbf{y}_{1:T}^P$ is also omitted because it consists of exact quantities.

The narrow $\pm\sigma$ interval around the mean values in Figures 7–9 shows that the BDLM is able to separate the raw datasets representing structural responses and external effects into its components.

On these figures, we can see how in approximately 5 months, the BDLM has gone from a broad prior for the hidden state to precise estimations of the true states.

Note that temperature and traffic loads are decomposed into their respective components not because they are themselves interesting, but because the frequency response depends on these components.

5.4.2 | Handling of missing data and outliers

This section presents how the BDLM is handling two common situations in SHM without requiring any modifications or special treatment. It first shows how the BDLM can fill the gap when structural responses and external effects data are missing. Figure 10 presents the filtered signal for the last 2 weeks of the dataset where 7 days of data are missing. The filtered signal is only marginally affected by missing data through an augmentation of the uncertainty in the AR process. Because $\phi^{B,AR} < 1$, when data is missing, the AR process tends to its zero-mean stationary state. When data is made available again, the estimation quickly returns to an accurate and precise estimation of the true hidden state.

Figure 11 presents the effect of having outliers corrupting the measured structural responses. In practical situations, outliers having unreasonable values are easy to identify and remove using thresholds.

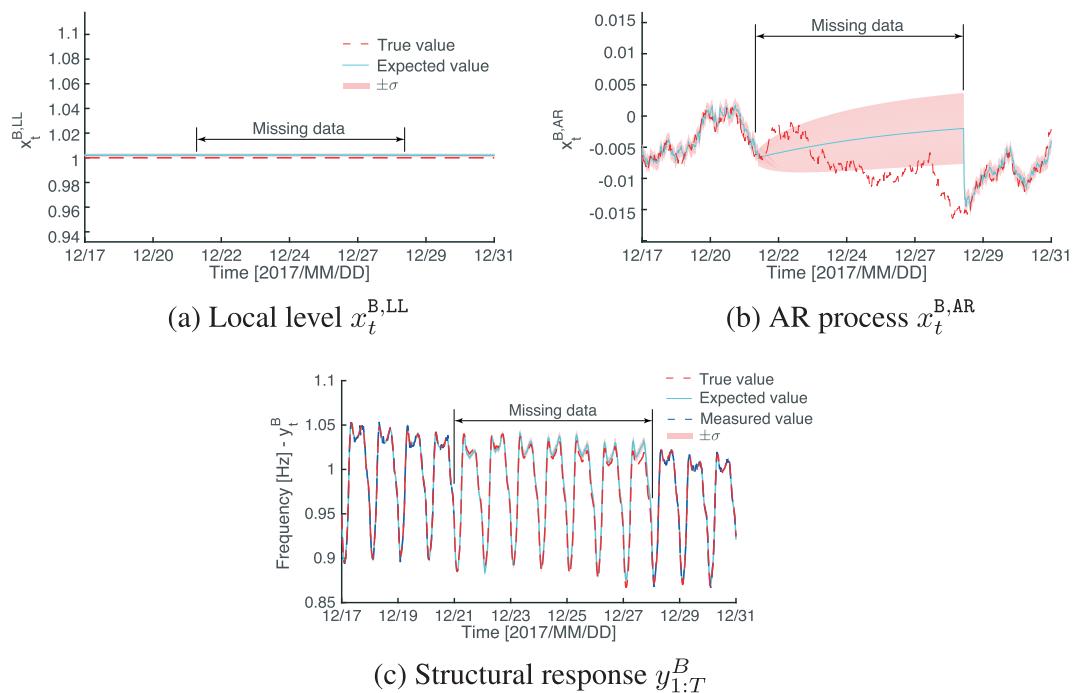


FIGURE 10 Example of results in the presence of missing data for the structural responses

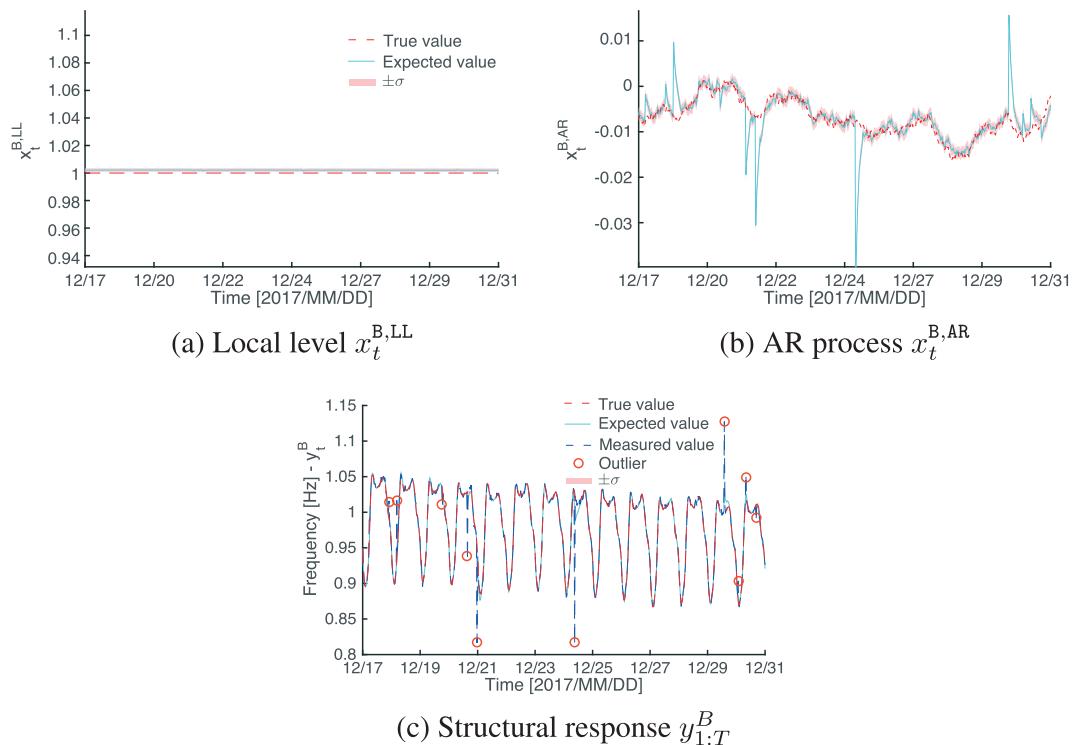


FIGURE 11 Example of results in the presence of outliers for the structural responses. In (c), outliers are indicated by a circle

Outliers that are in a gray-zone where they are at the same time unlikely, yet possible are the most problematic. BDLM is capable of handling those outliers in a seamless manner because filtering combines observations with a transition model. In this case, outliers introduce biases in the estimation of the AR process. Yet, the estimates of the local level representing the baseline behavior of the structure remains unaffected by the presence of outliers.

6 | CONCLUSION

This paper presents a framework for building, learning, and estimating BDLMs. Specifically, the contributions of this paper enable creating models of external effect and structural responses by superposing generic components, that is, baseline response, periodic cycles, autoregressive components, and regression components. The framework proposed offers robustness to numerical errors, outliers, and missing data. The EM calibration strategy has been shown to enable the quick estimation of unknown parameters. Moreover, it is compatible with nonuniform time steps that are common in practice.

Note that although the method proposed in this paper is able to handle multiple structural responses simultaneously, in practice, this would require special considerations regarding the definition of the model error covariance matrix. The study of this special case is beyond the scope of this paper.

The BDLM has the potential to enable a widespread application of SHM because (a) with minimal effort, it can be adapted to any type of structures, (b) it can process in real time, on a single desktop computer, the data acquired on thousands of structures and learn the model parameters. The BDLM presented in this paper is the first necessary step toward more advanced methods that will perform real-time autonomous anomaly detection.

ACKNOWLEDGEMENT

The author thanks Armen Der Kiureghian for his comments on the paper.

REFERENCES

- [1] L. Faravelli, S. Casciati, *Prog. Struct. Mater. Eng.* **2004**, 6(2), 104.
- [2] D. Bernal, D. Zonta, M. Pozzi, *Struct. Control Health Monit.* **2012**, 19(4), 535.
- [3] H. Zheng, A. Mita, *Struct. Control Health Monit.* **2008**, 15(7), 992.
- [4] F. Vicario, M. Q. Phan, R. Betti, R. W. Longman, *Struct. Control Health Monit.* **2015**, 22(5), 847.
- [5] E. M. Hernandez, *Struct. Control Health Monit.* **2013**, 20(4), 532.
- [6] H. Sohn, C. R. Farrar, N. F. Hunter, K. Worden, *Trans. Amer. Soc. Mech. Eng. J. Dyn. Syst. Meas. Control* **2001**, 123(4), 706.
- [7] G. Bartram, S. Mahadevan, *Struct. Control Health Monit.* **2013**.
- [8] C. R. Farrar, G. Park, D. W. Allen, M. D. Todd, *Struc. Control Health Monit.* **2006**, 13(1), 210.
- [9] S. Cho, R. K. Giles, B. F. Spencer, *Struct. Control Health Monit.* **2014**.
- [10] F. Casciati, L. Faravelli, *Smart Struct. Syst.* **2014**, 13(3), 343.
- [11] K. Yuen, *Bayesian Methods for Structural Dynamics and Civil Engineering*, Wiley, Singapore **2010**.
- [12] C. Farrar, K. Worden, *Structural Health Monitoring: A Machine Learning Perspective*, John Wiley & Sons, West Sussex, UK **2012**.
- [13] M. West, *in Honour of Adrian FM Smith* **2013**, 145.
- [14] M. West, J. Harrison, *Bayesian Forecasting and Dynamic Models*, Springer Series in Statistics, Springer, New York **1999**.
- [15] B. P. Gibbs, *Advanced Kalman Filtering, Least-Squares and Modeling: A Practical Handbook*, John Wiley & Sons, Hoboken, NJ, USA **2011**.
- [16] K. Murphy, *Machine Learning: A Probabilistic Perspective*, The MIT Press, Cambridge, MA, USA **2012**.
- [17] M. Chang, S. N. Pakzad, *J. Bridge Eng.* **2013**.
- [18] Y. Chen, M. Feng, *J. Eng. Mech.* **2009**, 135(4), 231.
- [19] P. Omenzetter, J. M. W. Brownjohn, *Smart Mater. Struct.* **2006**, 15(1), 129.
- [20] Z. Dzunic, J. G. Chen, H. Mobahi, O. Buyukozturk, J. W. Fisher III, *in Dynamics of Civil Structures*, vol. 2, Springer **2015**, 171.
- [21] A. Smyth, M. Wu, *Mech. Syst. Signal Proces.* **2007**, 21(2), 706.
- [22] H.-Q. Mu, K.-V. Yuen, *J. Eng. Mech.* **2015**, 141(1).
- [23] H. Sun, D. Feng, Y. Liu, M. Q Feng, *Comput.-Aided Civil Infrastruct. Eng.* **2015**, 30(11), 843.
- [24] I. Soljhell. **2009**, Bayesian forecasting and dynamic models applied to strain data from the göta river bridge, Master's Thesis.
- [25] Z. Ghahramani, G. E. Hinton, Parameter estimation for linear dynamical systems, *CRG-TR-96-2*, University of Toronto, Dept. of Computer Science **1996**.
- [26] P. Zarchan, H. Musoff, *Fundamentals of Kalman Filtering: A Practical Approach*, 208, Aiaa, Reston, VA, USA **2005**.
- [27] K. Mehrotra, P. R. Mahapatra, *IEEE Trans. Aerosp. Electron. Syst.* **1997**, 33(4), 1094.
- [28] K.-V. Yuen, S.-C. Kuok, *Earthq. Eng. Eng. Vibr.* **2010**, 9(2), 295.
- [29] R. Prado, M. West, *Time Series: Modeling, Computation, and Inference*, CRC Press, Boca Raton, FL, USA **2010**.
- [30] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, Morgan Kaufmann, San Mateo, USA **1988**.
- [31] G. Bierman, C. Thornton, *Automatica* **1977**, 13(1), 23.
- [32] D. Simon, *Optimal State Estimation: Kalman, H Infinity, and Nonlinear Approaches*, Wiley, Hoboken, NJ, USA **2006**.

- [33] K. P. Murphy, Switching kalman filters. Tachnical Report, Citeseer **1998**.
- [34] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Rubin, *Bayesian Data Analysis*, 3rd ed., CRC Press, Boca Raton, FL, USA **2014**.
- [35] K.-V. Yuen, S.-C. Kuok, *Mech. Syst. Signal Proces.* **2016**, *66*, 62.

How to cite this article: Goulet J-A. Bayesian Dynamic Linear Models for Structural Health Monitoring. *Struct Control Health Monit.* 2017;24:e2035. <https://doi.org/10.1002/stc.2035>

APPENDIX A

The complete vector of *state variables* (\mathbf{x}_t), the model *transition matrix* (\mathbf{A}_t), the model *observation matrix* (\mathbf{C}_t), the model *observation error covariance matrix* (\mathbf{R}_t), and the model *model error covariance matrix* (\mathbf{Q}_t) are presented below

$$\mathbf{x}_t = [x_t^{\text{B}, \text{LL}}, x_t^{\text{B}, \text{AR}}, x_t^{\text{T}, \text{LL}}, x_t^{\text{T1}, \text{S1}}, x_t^{\text{T1}, \text{S2}}, x_t^{\text{T2}, \text{S1}}, x_t^{\text{T2}, \text{S2}}, x_t^{\text{T}, \text{AR}}, x_t^{\text{T}, \text{LL}}, x_t^{\text{L}, \text{AR}}, x_t^{\text{P}}]^T$$

$$\mathbf{A}_t = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \phi^{\text{B}, \text{AR}} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \cos \omega^{\text{T1}} & \sin \omega^{\text{T1}} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -\sin \omega^{\text{T1}} & \cos \omega^{\text{T1}} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \cos \omega^{\text{T2}} & \sin \omega^{\text{T2}} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -\sin \omega^{\text{T2}} & \cos \omega^{\text{T2}} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \phi^{\text{T}, \text{AR}} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \phi^{\text{L}, \text{AR}} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\mathbf{C}_t = \begin{bmatrix} 1 & 1 & 0 & \phi^{\text{B}| \text{T}, \text{R}} & 0 & \phi^{\text{B}| \text{T}, \text{R}} & 0 & \phi^{\text{B}| \text{T}, \text{R}} & 0 & \phi^{\text{B}| \text{L}, \text{R}} & \phi^{\text{B}| \text{P}, \text{R}} \\ 0 & 0 & 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & \phi^{\text{L}| \text{P}, \text{R}} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\mathbf{R}_t = \begin{bmatrix} (\sigma_{\text{B}})^2 & 0 & 0 & 0 \\ 0 & (\sigma_{\text{T}})^2 & 0 & 0 \\ 0 & 0 & (\sigma_{\text{L}})^2 & 0 \\ 0 & 0 & 0 & \rightarrow 0 \end{bmatrix}$$

$$\mathbf{Q}_t = \text{blockdiag} (0, (\sigma^{\text{B}, \text{AR}})^2, 0, 0, 0, 0, 0, (\sigma^{\text{T}, \text{AR}})^2, 0, (\sigma^{\text{L}, \text{AR}})^2, \rightarrow \infty)$$