*Goulet, J.-A. and Koo, K. (2017). Empirical Validation of Bayesian Dynamic Linear Models in the Context of Structural Health Monitoring*
*Journal of Structural Engineering, ASCE.*

# Empirical Validation of Bayesian Dynamic Linear Models in the Context of Structural Health Monitoring

James-A. Goulet[1] and Ki Koo[2]

## Abstract

Bayesian Dynamic Linear Models (BDLM) are traditionally employed in the fields of applied statistics and Machine Learning. This paper performs an empirical validation of BDLM in the context of Structural Health Monitoring (SHM) for separating the observed response of a structure into sub-components. These sub-components describe the baseline response of the structure, the effect of traffic, and the effect of temperature. This utilization of BDLM for SHM is validated with data recorded on the Tamar Bridge (UK). This study is performed in the context of large-scale civil structures where missing data, outliers and non-uniform time steps are present. The study shows that the BDLM is able to separate observations into generic sub-components allowing to isolate the baseline behavior of the structure.
**Keywords**: Structural Health Monitoring (SHM), Bayesian, Dynamic Lineal Models, Kalman Filter, Bridge, infrastructure, Tamar Bridge

## INTRODUCTION

When developing a methodology for interpreting structural health monitoring data, working with simulated data is not equivalent to working with real data, also, controlled condition laboratory data is not equivalent to data acquired on full-scale structures. In real operation condition, it is common to have external effects that hinders the interpretation of data. Few studies are available where data interpretation methodologies have been validated for long-term applications in large-scale civil structures, where missing data and large outliers are common. Also, most studies found in the literature aims at detecting anomalies rather than modeling the structure behavior itself (Laory et al. 2011; Reynders et al. 2012; Farrar and Worden 2012; Balsamo et al. 2014). Authors such as Follen et al. (Follen et al. 2014) have sought to estimate the signature of a structure in order to detect anomalies through deviation from the normal response.

An alternative to existing approach is the Bayesian Dynamic Linear Models (*BDLM*), which takes its origin in the fields of Applied Statistics (Prado and West 2010; West 2013;

[1]Assistant Professor, Department of Civil, Geologic and Mining Engineering, Polytechnique Montreal, 2900 Edouard-Montpetit Montreal, Quebec, H3T 1J4 , CANADA (corresponding author). E-mail: James.A.Goulet@gmail.com

[2]Lecturer, Department of Engineering, University of Exeter, Prince of Wales Road Exeter, Devon UK, EX4 4SB, UK

West and Harrison 1999) and machine learning (Murphy 2012). More recently, Goulet (2017) has adapted the BDLM theory for the specificities of the SHM context. This perform an empirical validation of BDLM in the context of Structural Health Monitoring (SHM), for separating the observed response of a structure into sub-components. These sub-components describe the baseline response of the structure, the effect of traffic, and the effect of temperature. This sub-component identification is performed in the context of large-scale civil structures where missing data, outliers and non-uniform time steps are present.

This utilization of BDLM for SHM is validated in this paper with the data recorded on the Tamar Bridge (UK). Several researchers have studied the behavior of the Tamar Bridge since the start of monitoring campaign. Cross et al. (2011) initially created empirical models of the structure response using the linear regression technique. The dependent variables used in this paper were temperature and wind speed. The linear dependency between the temperature recorded at several locations was removed using Principal Component Analysis (Murphy 2012). Those authors found that temperature has a dominant effect over wind speed. In a subsequent study, Koo et al. (2013) confirmed that the wind speed has a negligible effect on the Bridge frequencies. Cross et al. (Cross et al. 2013) later created linear-regression-based empirical models using temperature, traffic loading and vertical accelerations. The conclusion of this study is that traffic loading is the dominant factor influencing the Tamar Bridge natural frequencies. Westgate and Brownjohn (2011) have developed a detailed finite element model for the structure. This model was used to further study the effect of temperature (Westgate et al. 2014a) and the effect of extreme loads (Westgate et al. 2014b) on the bridge behavior. The dependence between the structure frequencies and the traffic loading was confirmed by a study by Westgate et al. (2015) where empirical data were compared with predictions obtained from the structure finite element model.

## BAYESIAN DYNAMIC LINEAR MODELS

A Bayesian Dynamic Linear Model (BDLM) describes observations $\mathbf{y}_t$ at a time $t \in (1:T)$ by a superposition of hidden states $\mathbf{x}_t$. This superposition is formalized by the *observation model*

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

where $\mathbf{C}_t$ is the observation matrix indicating which hidden state contributes to observations and where $\mathbf{v}_t$ is the observation error. In BDLM, all quantities are described by Gaussian Random Variables. The *transition model* describes the temporal evolution of hidden states $\mathbf{x}_t$ between time steps $t-1$ and $t$ so that

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

$\mathbf{A}_t$ is the transition matrix defining changes between time steps and $\mathbf{Q}_t$ is transition model error. The key aspect is that the formulation associated with each hidden state $\mathbf{x}_t$ follows a generic form that depends on the type of component chosen. The main generic components useful for the context of SHM are;

**Local level:** baseline response of a structure

**Local trend/acceleration:** rate of change/acceleration in the baseline response
**Autoregressive:** time-dependent model prediction errors
**Periodic:** Sine-like phenomena such as daily or seasonal temperature cycles
**Regression:** dependence between hidden components associated with different observations.

The specific mathematical formulation associated with each component is detailed either by West and Harisson (1999) and by Goulet (2017). Components should be seen as building blocks which when assembled together, are able to model a wide variety of behaviour. The BDLM approach is modelling the response of structures using simple generic sub-components. Therefore, modelling errors remain and these errors are dependent over time steps. When the model is missing some physical phenomena and $\Delta t \to 0$, then model errors will be perfectly correlated because what is missing is the same between the two consecutive time steps; as $\Delta t$ increases, correlation is reduced. The autoregressive component is there to capture the time dependent model errors. One could see this component as a garbage bin where falls everything that cannot be explained by the local level/trend/acceleration, periodic or regression components. At the end the better the BDLM model is, the smaller is the amplitude of the AR state variables. Moreover, if by inspecting the AR state variable time-series estimates, one notice trends or patterns, it is an indication that wither (1) some explanatory sub-components are missing in the model or (2) the model calibration is inadequate.

No matter which generic component is employed, the hidden states are estimated using either the Kalman (KF) (Welch and Bishop 2001) or the UD Filter (Simon 2006). The KF and UD filter are both estimating $p(\mathbf{x}_t|\mathbf{y}_{1:t}) \sim N(\mathbf{x}_t; \boldsymbol{\mu}_{t|t}, \boldsymbol{\Sigma}_{t|t})$, where

$$(\boldsymbol{\mu}_{t|t}, \boldsymbol{\Sigma}_{t|t}) \quad = \quad \text{Filter}(\boldsymbol{\mu}_{t-1|t-1}, \boldsymbol{\Sigma}_{t-1|t-1}, \mathbf{A}_t, \mathbf{C}_t, \mathbf{Q}_t, \mathbf{R}_t). \tag{3}$$

Filtering is intended to be performed in an online procedure, i.e. as data is collected. On the other hand, the Kalman and UD smoother are intended to be employed offline in order to estimate $p(\mathbf{x}_t|\mathbf{y}_{1:T})$, where $T > t$. The Kalman filter and smoother are computationally more efficient than the UD algorithms. The UD algorithm performs a factorization of the covariance matrix so that it trades computational efficiency for an improved robustness toward numerical errors (Gibbs 2011).

In traditional application of the Kalman filter in SHM such as those performed by Chen and Feng (2009), Vicario et al. (2015), Mu and Yuen (2015) or Chang and Pakzad (2013), users must provide a dynamic model describing the structure behaviour. In contrast, with the BDLM approach, such model is not required because the structural response ant its dependency on external effects is model by generic sub-components.

## STRUCTURE AND DATASETS

The potential of the BDLM methodology is validated using data recorded on the Tamar Bridge, the UK. The structure is a suspension bridge with a main span of 335 m. The structure has gone through major renovation works that were completed in 2002 (Buckland 2003). Since 2007, natural frequencies of the bridge as well as external effects including temperature and traffic load are monitored. Details of the monitoring system and processing of the data are reported by Koo et al. (2013), Cross et al. (2013) and Brownjohn et al.(2007). The dataset

consists of the temperature and natural frequencies averaged over a period of 30 min and traffic load averaged over 60 min. The first year of data is employed as a training period to learn the parameters of the model and the rest of the data set is employed to test the model.

**Temperature**

Temperature data $y_t^T$ consists of averaged values over half-hour periods. Figure 1 shows the seasonal and daily temperature variation recorded on the bridge deck. In addition to daily and seasonal temperature cycle, Figure 1 shows that the daily cycle amplitude is larger during the summer than during the winter. This phenomenon is known as a *beat* (Roberts 2016) and can be represented by the interference between a 1 day and 365 days period components. Note that the blank space around the firsts months of 2009 indicates a long period with missing data. The left figure presents the entire dataset where the seasonal variability can be observed. The right figure presents only a 3 days period of data in order to display the daily variability.

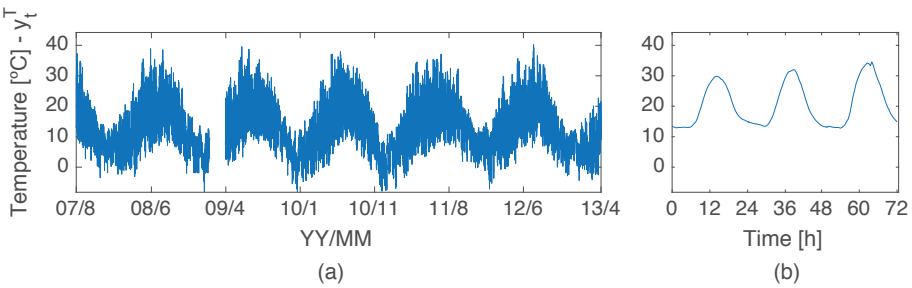



**Figure 1. Temperature data recorded on the bridge deck. (a) presents the entire dataset where the seasonal variability can be observed. (b) presents only a 3 days period of data in order to display the daily variability.**

**Traffic Load & Pattern**

The traffic load $y_t^L$ is estimated using the total traffic-load computed hourly from a toll booth, where vehicles are counted and separated by weight classes (Koo et al. 2013). Because traffic load is averaged hourly, it is assumed that the value for each full and half hours are the same. Figure 2 shows the total traffic load. The left figure presents the entire dataset where the Christmas holidays are identifiable by drops in the average traffic. The right figure presents only 3 days where the daily variation can be observed.

Note that no data is available beyond July 2011. In order to enable handling this situation where traffic data is missing, the Bayesian Dynamic Linear Model relies on the build-in probabilistic predictive estimate of the traffic load which itself depends on the average daily traffic pattern. The daily traffic patten is identified for two classes of days: *workdays* and *non-workdays*. Non-workdays are taken as weekends and the period between December 23$^{\text{rd}}$ and January 2$^{\text{nd}}$, i.e., Christmas holidays. Workdays are any other days. Figure 3 presents the daily traffic pattern obtained by averaging the traffic load for each day category during the training period. The traffic pattern in Figure 4 is normalized with to have a standard deviation equal to one and a mean of zero in order to be used as a regression variable.

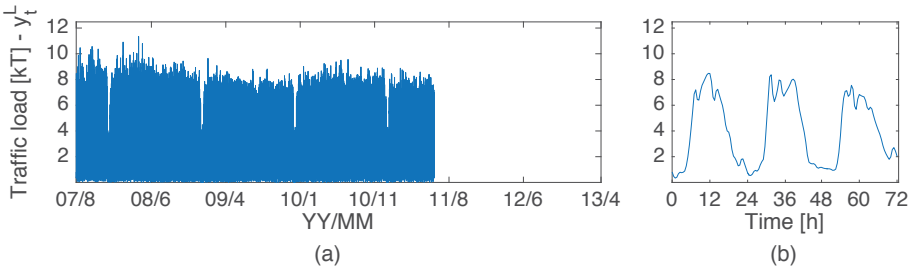



**Figure 2. Total traffic load in kilotons. (a) present the entire dataset where the Christmas holidays are identifiable by drops in the average traffic. (b) presents only 3 days where the daily variation can be observed.**

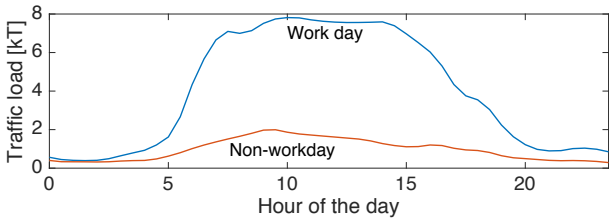



**Figure 3. Daily traffic patterns identified from the training period data for work-days and non-workdays.**

Figure 4 presents the traffic pattern $y_t^P$ for the entire time period corresponding to the dataset. The left figure presents the traffic pattern for the entire time period corresponding to the dataset. The right figure presents the traffic pattern for a period of 3 days. Note that on figures 2 and 4 the reduction in traffic intensity during the Christmas holidays is visible, however, it is not visible for weekends because of the density of the data presented.

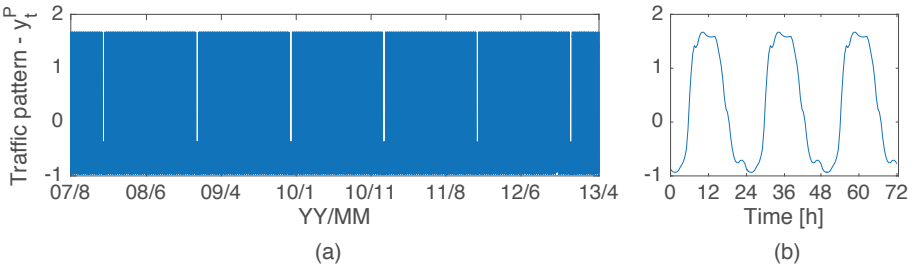



**Figure 4. (a) presents the traffic pattern for the entire time period corresponding to the dataset. (b) presents the traffic pattern for a period of 3 days.**

**Frequency**

For this study, only the first natural frequency $y_t^B$ is employed as characteristic responses of the structure. Frequencies were computed from acceleration responses using data-driven stochastic subspace identification (Peeters and De Roeck 1999). Hardware characteristics, sensor locations and data processing were reported by Koo et al. (2013). Figure 5 shows the evolution over time of the first natural frequency and blank segments represent missing data. The left figure presents the entire dataset and the right one presents only a period of three days. Note that the data is affected by a several outliers that display variability several times larger than normal values. The largest outliers occur around July 2009.

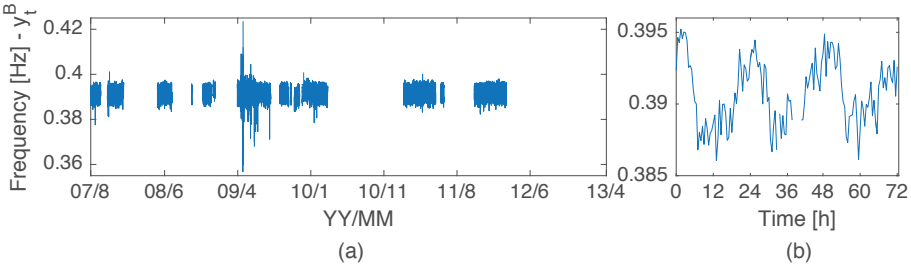



**Figure 5. Natural frequencies obtained from acceleration recordings. (a) presents the dataset and (b) one presents only a period of three days.**

**Missing Data & Non-uniform Time Steps**

When presenting the datasets for temperatures, traffic loads and frequencies, it was mentioned that missing data were present. However, due to the density of the data presented, it is only possible to see long-term missing data on Figures 1, 2 and 5. Figure 6 represents missing data at each time step by a cross. This representation shows that frequency is missing at several time steps because acceleration observations were missing. Yet, most of the missing data events have a short duration. Frequencies, temperature and traffic load, data is missing in respectively 65%, 4% and 31% of the time steps. Overall at least one data is missing in 71% of the time steps. This high number of missing data underlines the importance of having a method capable of handling them.

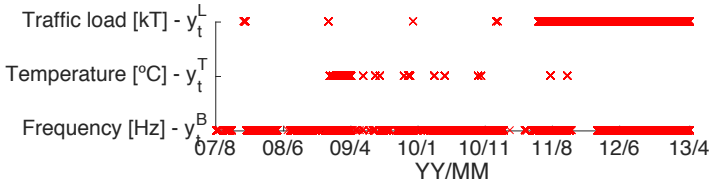



**Figure 6. Each cross represents missing data at a time step for an observation. Frequencies, temperature and traffic load, data is missing in respectively 65%, 4% and 31% of the time steps.**

In addition to missing data, another challenge is the non-uniformity of time steps as presented in Figure 7. 99.98% of the time steps have a duration of 30 min ($\Delta t = \frac{1}{48}$), however, in some exception, the time steps between recordings is around 24h and in one case 5 days.

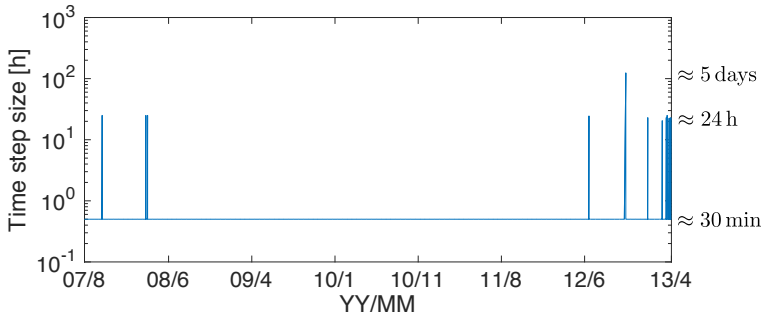



**Figure 7. Time step size is presented on a log-scale for the entire dataset duration.**

## BAYESIAN DYNAMIC LINEAR MODEL CONSTRUCTION

This section describes how to construct the Bayesian Dynamic Linear model for the Tamar Bridge. The superscript nomenclature in this paper follows: for observations, ($^{\mathtt{T}}$) temperature, ($^{\mathtt{P}}$) traffic pattern, ($^{\mathtt{L}}$) traffic load and ($^{\mathtt{B}}$) frequencies and for hidden state variables, ($^{\mathtt{LL}}$) local level, ($^{\mathtt{S}}$) cyclic component, ($^{\mathtt{AR}}$) autoregressive component, and ($^{\mathtt{R}}$) regression component. Note that Local trend and Local acceleration components are not employed in this case-study. The global matrices defining the Bayesian Dynamic Linear model are

$$\begin{aligned}
\mathbf{y}_t &= [y_t^{\mathtt{B}}, y_t^{\mathtt{T}}, y_t^{\mathtt{L}}, y_t^{\mathtt{P}}]^\intercal \\
\mathbf{x}_t &= [\mathbf{x}_t^{\mathtt{B}}, \mathbf{x}_t^{\mathtt{T}}, \mathbf{x}_t^{\mathtt{L}}, x_t^{\mathtt{P}}]^\intercal \\
\mathbf{A}_t &= \text{block diag}\left(\mathbf{A}_t^{\mathtt{B}}, \mathbf{A}_t^{\mathtt{T}}, \mathbf{A}_t^{\mathtt{L}}, A_t^{\mathtt{P}}\right) \\
\mathbf{C}_t &= \text{block diag}\left(\mathbf{C}_t^{\mathtt{B}}, \mathbf{C}_t^{\mathtt{T}}, \mathbf{C}_t^{\mathtt{L}}, C_t^{\mathtt{P}}\right) \\
\mathbf{R}_t &= \text{block diag}\left(R_t^{\mathtt{B}}, R_t^{\mathtt{T}}, R_t^{\mathtt{L}}, R_t^{\mathtt{P}}\right) \\
\mathbf{Q}_t &= \text{block diag}\left(\mathbf{Q}_t^{\mathtt{B}}, \mathbf{Q}_t^{\mathtt{T}}, \mathbf{Q}_t^{\mathtt{L}}, Q_t^{\mathtt{P}}\right)
\end{aligned} \tag{4}$$

where each component of these model matrices are defined below. Note that the ordering of variable does not influence results. Model parameters to be estimated are regrouped in the set

$$\begin{aligned}
\mathcal{P} = \{&\overbrace{\phi^{\mathtt{B}|\mathtt{L},\mathtt{R}}, \phi^{\mathtt{B}|\mathtt{T},\mathtt{R}}, \phi^{\mathtt{B}|\mathtt{P},\mathtt{R}}, \phi^{\mathtt{L}|\mathtt{P},\mathtt{R}}}^{\text{regression coefficients}}, \overbrace{\phi^{\mathtt{B},\mathtt{AR}}, \phi^{\mathtt{T},\mathtt{AR}}, \phi^{\mathtt{L},\mathtt{AR}}}^{\text{autocorr. coefficients}}, \dots \\
&\underbrace{\sigma^{\mathtt{B},\mathtt{LL}}, \sigma^{\mathtt{L},\mathtt{LL}}, \sigma^{\mathtt{T},\mathtt{LL}}, \sigma^{\mathtt{P},\mathtt{LL}}}_{\sqrt{\text{local level variance}}}, \underbrace{\sigma^{\mathtt{B},\mathtt{AR}}, \sigma^{\mathtt{L},\mathtt{AR}}, \sigma^{\mathtt{T},\mathtt{AR}}}_{\sqrt{\text{autocorr. variance}}}, \underbrace{\sigma^{\mathtt{B}}, \sigma^{\mathtt{L}}, \sigma^{\mathtt{T}}, \sigma^{\mathtt{P}}}_{\sqrt{\text{meas. variance}}}\}
\end{aligned} \tag{5}$$

where $\phi^{\cdot|\cdot,\mathtt{R}}$ describe regression coefficient, $\phi^{\cdot,\mathtt{AR}}$ describe autoregression coefficients and $\phi^{\cdot,\cdot}$ describe standard deviations.

**Temperature –** $y_t^{\mathtt{T}}$

There are five components involved in the temperature model: (1) a local level, (2-3) two periodic signals with a period of one day ($\omega^{\mathtt{T1}} = 2\pi\Delta t$) and 365 days ($\omega^{\mathtt{T2}} = \frac{2\pi}{365}\Delta t$), (4) one beat periodic signal with a period of 1.0027 days ($\omega^{\mathtt{T3}} = (2\pi - \frac{2\pi}{365}\Delta t$) for modeling the change in temperature amplitude between summers and winters, and (5) an autoregressive process of order 1. The vector of hidden state variables for the temperature is

$$\mathbf{x}_t^{\mathtt{T}} = [\ \underbrace{x_t^{\mathtt{T,LL}}}_{\text{local level}}, \underbrace{x_t^{\mathtt{T1,S1}}, x_t^{\mathtt{T1,S2}}}_{\text{cycle, } p=1\text{ day}}, \underbrace{x_t^{\mathtt{T2,S1}}, x_t^{\mathtt{T2,S2}}}_{\text{cycle, } p=365\text{ day}}, \underbrace{x_t^{\mathtt{T3,S1}}, x_t^{\mathtt{T3,S2}}}_{\text{cycle, } p=48.2\text{ day}}, \underbrace{x_t^{\mathtt{T,AR}}}_{\text{AR process}}\ ] \tag{6}$$

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

$$\mathbf{C}_t^{\mathtt{T}} = [1, 1, 0, 1, 0, 1, 0, 1] \tag{7}$$

where zeros indicate that hidden state variables $x_t^{\mathtt{T1,S2}}$, $x_t^{\mathtt{T2,S2}}$ and $x_t^{\mathtt{T3,S2}}$ do not contribute to the temperature observations. The measurement noise variance is $R_t^{\mathtt{T}} = (\sigma^{\mathtt{T}})^2$. The transition matrix is

$$\mathbf{A}_t^{\mathtt{T}} = \text{block diag}\left(1, \begin{bmatrix} \cos\omega^{\mathtt{T1}} & \sin\omega^{\mathtt{T1}} \\ -\sin\omega^{\mathtt{T1}} & \cos\omega^{\mathtt{T1}} \end{bmatrix}, \begin{bmatrix} \cos\omega^{\mathtt{T2}} & \sin\omega^{\mathtt{T2}} \\ -\sin\omega^{\mathtt{T2}} & \cos\omega^{\mathtt{T2}} \end{bmatrix}, \begin{bmatrix} \cos\omega^{\mathtt{T3}} & \sin\omega^{\mathtt{T3}} \\ -\sin\omega^{\mathtt{T3}} & \cos\omega^{\mathtt{T3}} \end{bmatrix}, \phi^{\mathtt{T,AR}}\right). \tag{8}$$

Block matrices with sine and cosine components are defined using the generic formulation for periodic components presented by Goulet (2017) and West and Harisson (1999). The model error covariance is

$$\mathbf{Q}_t^{\mathtt{T}} = \text{block diag}\left((\sigma^{\mathtt{T,LL}})^2, 0, 0, 0, (\sigma^{\mathtt{T,AR}})^2\right). \tag{9}$$

**Traffic Pattern –** $y_t^{\mathtt{P}}$

Because the traffic pattern is a regressor variable, it is described by a single local level component. Matrices defining this component are

$$\mathbf{x}_t^{\mathtt{P}} = x_t^{\mathtt{P}},\ A_t^{\mathtt{P}} = 1, C_t^{\mathtt{P}} = 1,\ R_t^{\mathtt{P}} = (\sigma^{\mathtt{P}})^2,\ Q_t^{\mathtt{P}} = (\sigma^{\mathtt{P,LL}})^2. \tag{10}$$

**Traffic Load –** $y_t^{\mathtt{L}}$

There are two independent components involved in the traffic load: (1) a local level and (2) an autoregressive process of order 1. 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

$$\mathbf{x}_t^{\mathtt{L}} = [\ \underbrace{x_t^{\mathtt{L,LL}}}_{\text{local level}}\ ,\ \underbrace{x_t^{\mathtt{L,AR}}}_{\text{AR process}}\ ] \tag{11}$$

where the ordering of each component remains the same for other matrices $\mathbf{C}_t^{\mathtt{L}}$, $\mathbf{A}_t^{\mathtt{L}}$ and $\mathbf{Q}_t^{\mathtt{L}}$. The observation matrix is $\mathbf{C}_t^{\mathtt{L}} = [1, 1]$, the measurement noise variance is $R_t^{\mathtt{L}} = (\sigma_{\mathtt{L}})^2$, and the transition matrix is $\mathbf{A}_t^{\mathtt{L}} = \text{block diag}\left(1, \phi^{\mathtt{L,AR}}\right)$ and the model error covariance is $\mathbf{Q}_t^{\mathtt{L}} = \text{block diag}\left((\sigma^{\mathtt{L,LL}})^2, (\sigma^{\mathtt{L,AR}})^2\right)$. Note that here $R_t^{\mathtt{L}}$ accounts for uncertainty in the average traffic. The dependent component is taken into account by introducing the regression coefficient in the global observation matrix

$$[\mathbf{C}_t]_{3,13} = \phi^{\mathtt{L|P,R}}. \tag{12}$$

The regression component ensures that the effect of the traffic pattern is included in the traffic load observation. Note that the position $(3, 13)$ in the global observation matrix depends on the ordering chosen for state variables.

**Frequency –** $y_t^{\mathtt{B}}$

There are two independent components involved in the frequency: (1) a local level and (2) an autoregressive process of order 1. There are also six dependent components: (3-9) the six regression components describe the dependency between the frequency and temperature or traffic components. The vector of state variables for the frequency is

$$\mathbf{x}_t^{\mathtt{B}} = [\ \underbrace{x_t^{\mathtt{B,LL}}}_{\text{local level}}\ ,\ \underbrace{x_t^{\mathtt{B,AR}}}_{\text{AR process}}\ ] \tag{13}$$

where the ordering of each component remains the same for other matrices $\mathbf{C}_t^{\mathtt{B}}$, $\mathbf{A}_t^{\mathtt{B}}$ and $\mathbf{Q}_t^{\mathtt{B}}$. The observation matrix is $\mathbf{C}_t^{\mathtt{B}} = [1, 1]$, the measurement noise variance is $R_t^{\mathtt{B}} = (\sigma_{\mathtt{B}})^2$, and the transition matrix is $\mathbf{A}_t^{\mathtt{B}} = \text{block diag}\left(1, \phi^{\mathtt{B,AR}}\right)$ and the model error covariance is $\mathbf{Q}_t^{\mathtt{B}} = \text{block diag}\left((\sigma^{\mathtt{T,LL}})^2, (\sigma^{\mathtt{B,AR}})^2\right)$. The dependent components are taken into account by introducing the regression coefficient in the global observation matrix

$$[\mathbf{C}_t]_{1,13} = \phi^{\mathtt{B|P,R}},\ \ [\mathbf{C}_t]_{1,12} = \phi^{\mathtt{B|L,R}},\ \ [\mathbf{C}_t]_{1,4} = [\mathbf{C}_t]_{1,6} = [\mathbf{C}_t]_{1,8} = [\mathbf{C}_t]_{1,9} = \phi^{\mathtt{B|T,R}} \tag{14}$$

so that the effect of temperature and traffic pattern is included in the frequency observation. Figure 8 presents one time slice from the Dynamic Bayesian Network (i.e. graphical model) describing the BDLM.

**Model Parameter Estimation**

Unknown model parameters $\mathcal{P}$ can typically be estimated with a training period of six months. In this case, because of the large amount of missing data, the training period was set to one year. For all hidden variables except those related to autoregressive components, the initial state $\mathbf{x}_0$ is described by a broad (i.e. non-informative) prior where each component has a mean of $[\boldsymbol{\mu}_{\mathbf{0}}]_i = 0, \forall i$ and a covariance $[\boldsymbol{\Sigma}_{\mathbf{0}}]_{ii} = 10^6, \forall i,\ [\boldsymbol{\Sigma}_{\mathbf{0}}]_{ij} = 0, \forall i \neq j$. For the autoregressive components, the initial state is defined as the stationary distribution of the AR process, i.e.$[\boldsymbol{\mu}_{\mathbf{0}}]_i = 0,\ [\boldsymbol{\Sigma}_{\mathbf{0}}]_{ii} = \frac{\sigma^{\mathtt{AR}}}{(1-\phi^{\mathtt{AR}})^2}$.

Optimal values $\mathcal{P}^*$ for unknown parameters are estimated using the parameter-wise EM algorithm because the matrix-wise approach was repeatedly caught in local maxima (Goulet 2017). The parameters values maximizing the log-likelihood are:

$$\begin{aligned}\mathcal{P}^* = \{&\overbrace{-2.8\times10^{-4}, -2.4\times10^{-5}, -2.0\times10^{-3}, 2.8}^{\text{regression coefficients}},\ \overbrace{0.91, 0.99, 0.91s}^{\text{autocorr. coefficients}}, \ldots \\ &\underbrace{2.4\times10^{-6}, 4\times10^{-7}, 9\times10^{-7}, 0.14}_{\sqrt{\text{local level variance}}},\ \underbrace{5\times10^{-4}, 0.28, 0.56}_{\sqrt{\text{autocorr. variance}}},\ \ldots \\ &\underbrace{9\times10^{-4}, 5\times10^{-5}, 2\times10^{-5}, 2\times20^{-7}}_{\sqrt{\text{meas. variance}}}\}\end{aligned} \tag{15}$$

where the ordering of parameters is identical as in Equation 5 and each unit is consistant with observations. The stopping criterion for the model parameter optimization algorithm is when

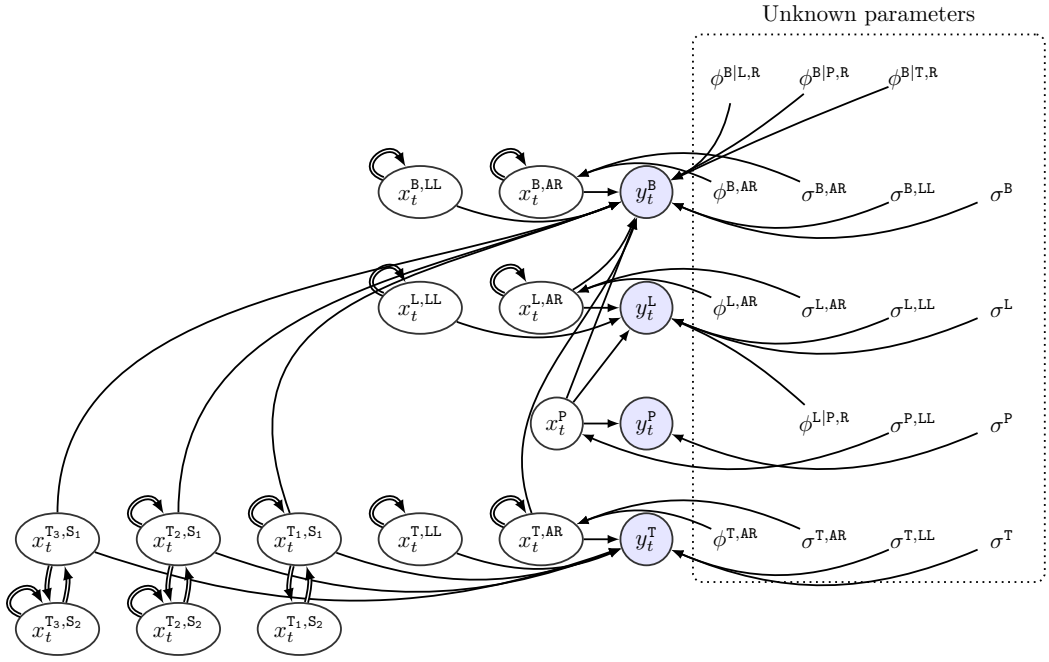



**Figure 8. Graphical model (i.e. Dynamic Bayesian network) representing the causal dependencies between each component of the model. Circles represent random variables; links correspond to causal relations; double-line links are a shorthand notation for links between time steps. Nodes without a border represent deterministic parameters; color-filled nodes correspond to *observed variables*; white-filled nodes correspond to *unobserved/hidden variables*.**

the change in the log-likelihood for 18 (i.e. the number of unknown parameters) subsequent loops is less than $10^{-3}\%$ of the likelihood. For the parameter set $\mathcal{P}^*$, the log-likelihood of the data in the training set is 12767. For indication, the root-mean-square error between predicted and observed frequencies is equal to $9.97 \times 10^{-4}$ Hz.

The optimal regression coefficient values identified indicate that the effect on frequencies of traffic loading dominates the effect of temperature. By multiplying the respective regression coefficient ($\phi^{B|P,R}$ and $\phi^{B|T,R}$) by the standard deviation of their associated component time-series, we find that traffic loading has a relative importance more than 20 times higher than the temperature. Although the optimal regression coefficient for temperature is larger than zero, the contribution of temperature on the first natural frequency is negligible compared to traffic loading.

## MODEL STATE ESTIMATION

The Bayesian Dynamic Linear Model uses the set of parameters $\mathcal{P}^*$ to decompose each observation in its hidden components as identified in Section 4. For figures presented in this section, all left graphs present the entire dataset and the right graphs present two weeks of data. Note that the vertical scale of each left and right graph is identical. The solid cyan line

represents the estimated expected value and the shaded region the $\pm 1\sigma$ region. The dashed red line represents the observed data. Note that in terms of computational requirements, estimating the BDLM for $6\frac{1}{2}$ years of data ($\approx 98000$ time steps), takes approximately one minute for the filtering algorithm and twice as much for the smoothing algorithm (note: calculations were performed on a laptop computer).

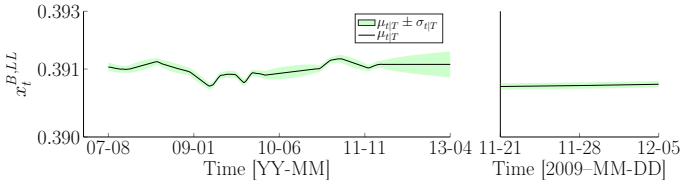
Figure 9 presents the smoothed estimates for components related to the structure frequency. Figure 9a presents the local level which corresponds to the baseline behavior of the structure without the effect of loading and temperature. The local level in Figure 9a displays a subtle non-stationary behavior where the amplitude of the change over the entire dataset is approximately $0.15\%$ of the average frequency. Note that the BDLM does not provide indication allowing to identify the cause of these changes. However, from a structural engineering standpoint, because this non-stationarity is also non-monotonic, it is unlikely to be caused by a change in the condition of the structure.

The autoregression component in Figure 9b displays, as expected, a stationary behavior. The structure frequency in Figure 9c shows that the model predictions are in agreement with observations. Also, the pattern of the predictions when data is missing is compatible with neighbouring data. Note that the large outliers in frequency observations visible in Figure 5 have not affected the identification of the baseline behavior presented in Figure 9a. Finally, large spikes around $5 \times 10^{-3}$ in the $\pm 1\sigma$ region for Figures 9b & c are caused by the combined effect of missing data and time steps that are significantly longer than normal.

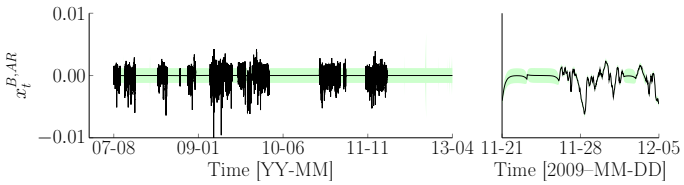
Figure 10 presents the smoothed estimates related to the temperature. Figures 10a shows that the average temperature is stationary across the entire dataset. Figures 10b & c are of particular interest because it shows the BDLM can identify daily and seasonal temperature cycles. Figures 10f shows that the model is in good agreement with the data and that it can provide temperature estimates when data is missing. Figure 11 presents the smoothed estimates related to the traffic load. Figure 11a shows that the average traffic load is stationary over the course of the monitoring period. On Figure 11c, the BDLM predictions enable to continue estimating other components in the model even after the first half of 2011 where traffic data is missing. Note that temperature and traffic load observations are divided in sub-components in order to model the dependency between the frequency observations and these sub-components.

## CONCLUSION

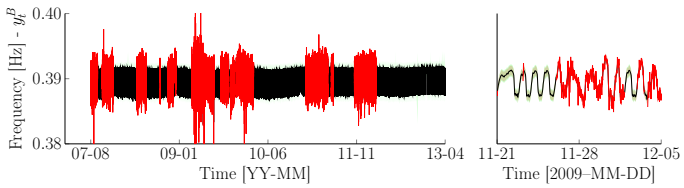
The results presented in this paper confirm that the Bayesian Dynamic Linear Model is applicable for SHM in the context of full-scale structures. The example shows that the BDLM is able to separate observations into generic sub-components allowing to isolate the baseline behavior of the structure. Moreover, the example demonstrates that the BDLM can operate in the context where outliers, missing data and non-uniform time steps are common. Also, because it takes approximately a minute to process $6\frac{1}{2}$ years of data ($\approx 98000$ time steps), the model could eventually be employed for the treatment in real-time of data coming from dozens of sensors located on thousands of structures. The BDLM model presented in this paper is able to model the response of a structure with respect to change in its environment. This is the first necessary step toward new methodologies capable of detecting changes in the baseline behavior of structures.




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




(b) AR process $x_t^{\mathtt{B,AR}}$




(c) Structural response $y_t^B$

**Figure 9. Illustration of the model component separation and smoothed predictions for the structural response.**

## ACKNOWLEDGEMENTS

The first author thanks the *Swiss National Science Foundation*, the *Fonds de recherche du Qubec  Nature et technologies (FRQNT)*, and the *National Research Council of Canada* (RGPIN-2016-06405) for funding this research. Tamar Bridge data were available via EPSRC grant EP/F035401/1.

## References

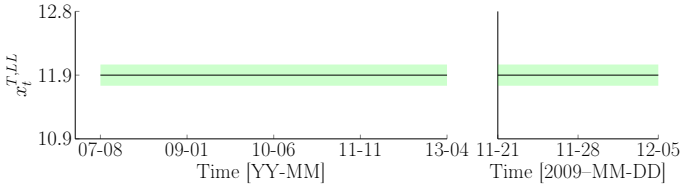
Balsamo, L., Mukhopadhyay, S., Betti, R., and Lus, H. (2014). "Damage detection using flexibility proportional coordinate modal assurance criterion." *Topics in Modal Analysis*, Vol. 7, Springer, 1–8.

Brownjohn, J., Pavic, A., Carden, P., and Middleton, C. (2007). "Modal testing of Tamar suspension bridge." *Proceedings of the IMAC XXV conference*, Orlando, USA, 19–22.

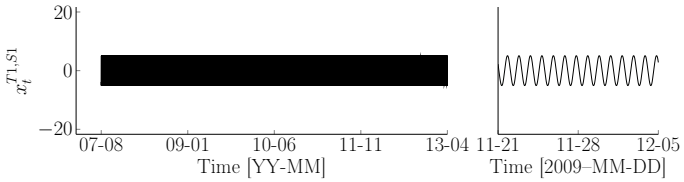
Buckland, P. G. (2003). "Increasing the load capacity of suspension bridges." *Journal of Bridge Engineering*, 8(5), 288–296.

Chang, M. and Pakzad, S. N. (2013). "Observer kalman filter identification for output-only systems using interactive structural modal identification toolsuite." *Journal of Bridge Engineering*.

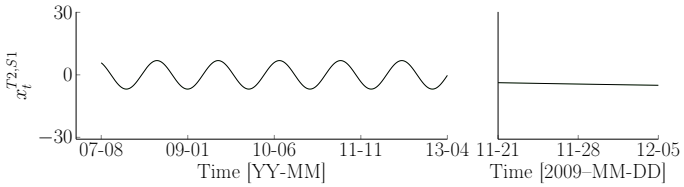
Chen, Y. and Feng, M. (2009). "Structural health monitoring by recursive Bayesian filtering." *Journal of Engineering Mechanics*, 135(4), 231–242.



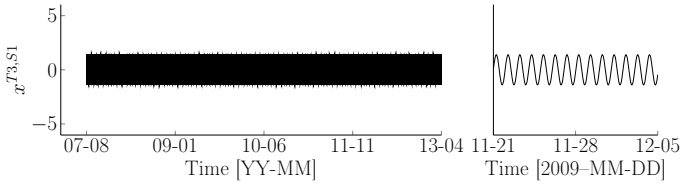

(a) Local level $x_t^{\mathtt{T,LL}}$



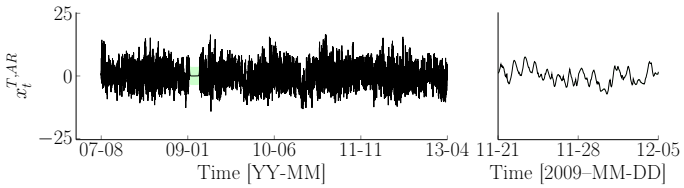

(b) Daily cycle $x_t^{\mathtt{T1,S1}}$



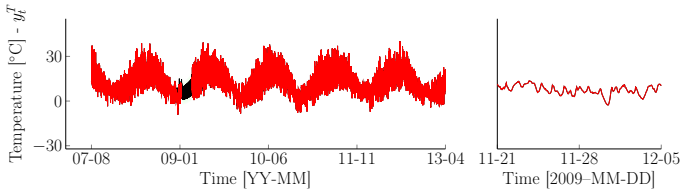

(c) Seasonal cycle $x_t^{\mathtt{T2,S1}}$




(d) Beat cycle $x_t^{\mathtt{T3,S1}}$

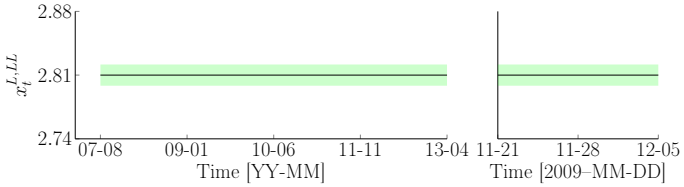



(e) AR process $x_t^{\mathtt{T,AR}}$

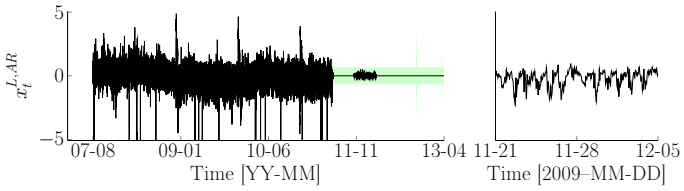



(f) Temperature $y_t^T$

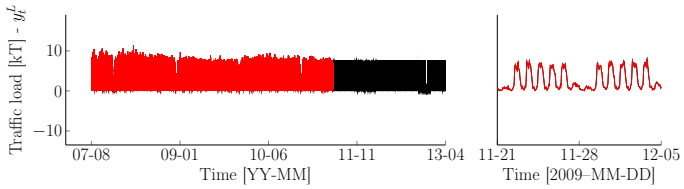
**Figure 10. Illustration of the model component separation and smoothed predictions for the temperature.**




(a) Local level $x_t^{\mathrm{L,LL}}$




(b) AR process $x_t^{\mathrm{L,AR}}$




(c) Traffic load $y_t^L$

**Figure 11. Illustration of the model component separation and smoothed predictions for the traffic load.**

Cross, E., Koo, K., Brownjohn, J., and Worden, K. (2013). "Long-term monitoring and data analysis of the tamar bridge." *Mechanical Systems and Signal Processing*, 35(1–2), 16–34.

Cross, E., Worden, K., Koo, K. Y., and Brownjohn, J. M. (2011). "Modelling environmental effects on the dynamic characteristics of the tamar suspension bridge." *Dynamics of Bridges, Volume 5*, Springer, 21–32.

Farrar, C. and Worden, K. (2012). *Structural Health Monitoring: A Machine Learning Perspective*. John Wiley & Sons.

Follen, C. W., Sanayei, M., Brenner, B. R., and Vogel, R. M. (2014). "Statistical bridge signatures." *Journal of Bridge Engineering*.

Gibbs, B. P. (2011). *Advanced Kalman filtering, least-squares and modeling: a practical handbook*. John Wiley & Sons.

Goulet, J. (2017). "Bayesian dynamic linear models for structural health monitoring." *Structural Control and Health Monitoring*, e2035. https://doi.org/10.1002/stc.2035.

Koo, K., Brownjohn, J., List, D., and Cole, R. (2013). "Structural health monitoring of the Tamar suspension bridge." *Structural Control and Health Monitoring*, 20(4), 609–625.

Laory, I., Trinh, T., and Smith, I. (2011). "Evaluating two model-free data interpretation methods for measurements that are influenced by temperature." *Advanced Engineering Informatics*, 25(3), 495–506.

Mu, H.-Q. and Yuen, K.-V. (2015). "Novel outlier-resistant extended kalman filter for robust online structural identification." *Journal of Engineering Mechanics*, 141(1), DOI: 10.1061/(ASCE)EM.1943–7889.0000810.

Murphy, K. (2012). *Machine learning: a probabilistic perspective*. The MIT Press.

Peeters, B. and De Roeck, G. (1999). "Reference-based stochastic subspace identification for output-only modal analysis." *Mechanical systems and signal processing*, 13(6), 855–878.

Prado, R. and West, M. (2010). *Time series: modeling, computation, and inference*. CRC Press.

Reynders, E., Wursten, G., and De Roeck, G. (2012). "Nonlinear system identification for vibration-based structural health monitoring." *Proceedings of NCTAM 2012*, 1–7.

Roberts, G. E. (2016). *From Music to Mathematics: Exploring the Connections*. JHU Press.

Simon, D. (2006). *Optimal state estimation: Kalman, H infinity, and nonlinear approaches*. Wiley.

Vicario, F., Phan, M. Q., Betti, R., and Longman, R. W. (2015). "Output-only observer/kalman filter identification (o3kid)." *Structural Control and Health Monitoring*, 22(5), 847–872.

Welch, G. and Bishop, G. (2001). *An introduction to the Kalman filter*. Chapel Hill, USA.

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

West, M. and Harrison, J. (1999). *Bayesian Forecasting and Dynamic Models*. Springer Series in Statistics. Springer New York.

Westgate, R. and Brownjohn, J. (2011). "Development of a Tamar Bridge finite element model." *Conference Proceedings of the Society for Experimental Mechanics Series 3*, Vol. 5, Springer, 13–20.

Westgate, R., Koo, K.-Y., and Brownjohn, J. (2014a). "Effect of solar radiation on suspension bridge performance." *Journal of Bridge Engineering*, 20(5), 04014077.

Westgate, R., Koo, K.-Y., and Brownjohn, J. (2015). "Effect of vehicular loading on suspension bridge dynamic properties." *Structure and Infrastructure Engineering*, 11(2), 129–144.

Westgate, R., Koo, K.-Y., Brownjohn, J., and List, D. (2014b). "Suspension bridge response due to extreme vehicle loads." *Structure and Infrastructure Engineering*, 10(6), 821–833.