

INTRODUCING DATA-CENTRIC ENGINEERING TO INSTRUMENTED INFRASTRUCTURE

L.J. Butler^{1,5}, D-H. Lau^{2,5}, A. Gregory^{2,5}, M. Girolami^{4,5} and M.Z.E.B. Elshafie³*

¹*Lassonde School of Engineering, York University, Toronto, Canada*

²*Department of Statistics, Imperial College London, London, United Kingdom*

³*Department of Civil Engineering, Qatar University, Doha, Qatar*

⁴*Department of Engineering, University of Cambridge, Cambridge, United Kingdom*

⁵*Lloyd's Register Foundation Programme on Data-Centric Engineering, The Alan Turing Institute, London, United Kingdom*

** Corresponding author*

ABSTRACT A variety of methods exist in the structural health monitoring literature that aim to combine the observed data and predicted outputs from physics-based models (e.g. model updating and calibration). Typically implemented on a case-by-case basis, there is currently no unifying procedure or method for formally synthesising this combination of information. Data-centric engineering (DCE) is an emerging class of analytical approaches that is aimed at studying engineered systems and assets through the synthesis of various data-driven and physics-based models. DCE may also involve procedures for fusing sensor data from both experimental and operational systems. This paper introduces a DCE-based approach using data collected from experimental and operational railway structures (bridges and sleepers), which have been instrumented with advanced fibre optic sensors (FOS). This study provides a high-level definition of DCE approaches to studying instrumented infrastructure and discusses some of the challenges in implementing these methods in practice. An example of an implementation of a Gaussian-process based DCE method is provided, with the goal of predicting the response of operational instrumented rail infrastructure (i.e. concrete sleepers) over time. Some practical issues of this implementation relate to the measurement systems, data acquisition rate, efficiently batch processing the data and accounting for uncertainty in the response predictions. In leveraging both the information gained from real-time measurement data and from traditional analytics or physics-based methods, a DCE-based modelling approach can provide unique a set information and insights into the operational performance of infrastructure.

1. Introduction

Newly constructed infrastructure systems and those which have been in operation for over 100 years, underpin both the social and economic well-being of societies around the world. The design, construction, operation, maintenance and eventual decommissioning phase form a life-cycle chain within which each link plays a crucial role in determining the whole life performance (and value) of an infrastructure asset.

There are a number of challenges facing the infrastructure management sector. In particular, these include pressures to maintain adequate levels of asset performance under shrinking financial budgets; aging assets which require ever-increasing maintenance; tighter government regulations requiring greater accountability and risk assessment; and the unknown future effects of climate change (ASCE, 2017; CSCE 2016). Gathering of reliable operational data is the crucial link between assessing performance and decision-making – it enables infrastructure to function at its maximum efficiency (Bowers et al. 2016). With the advent of new sensing technologies which can record performance data and be readily integrated or retrofitted within our built environment, traditional civil engineering practice has added a powerful

toolkit to its arsenal. Once acquired, this data can provide the information to undertake ongoing condition monitoring, be used for damage detection and to form the basis for predicting future structural health.

Typically, engineers have utilised physics-based or numerical models in order to simulate (and predict) the performance of the variety of infrastructure assets that they are tasked with designing or assessing. Under more complex design and assessment requirements, experimental testing and simulation are often used to provide additional information (e.g. material properties, boundary conditions, loading scenarios, etc.) to improve the accuracy of the numerical (physics-based) model. Measurement data obtained from operational structures provides yet a third source of information for better understanding infrastructure assets. However, as sensing systems become more advanced and more pervasive across the built environment, the generation of larger and more complex data sets is inevitable. Developing robust and portable techniques for processing, managing and interpreting these vast data sets is crucial for realizing the full benefits of long-term condition monitoring.

As outlined in Figure 1, there are several potential approaches for modelling engineered systems. These include model-driven approaches (e.g. closed-form physics-based, FE models, etc.), data-driven approaches (i.e. model free or those considering only information gathered from performance data), model-updating or system identification approaches (i.e. those seeking to infer critical system parameters based on response of the system), and a new class of approaches based on data-centric engineering.

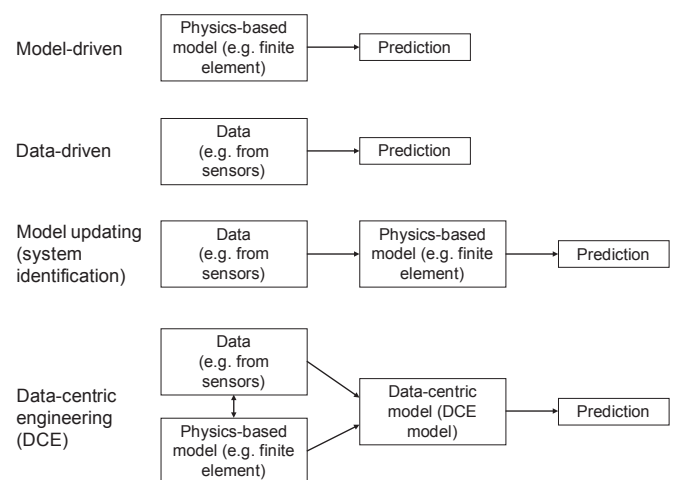
A variety of data-driven approaches have emerged as the primary means for developing performance models and for detecting statistically significant trends which may indicate deterioration or damage (Farrar and Worden 2013). Statistical techniques for data reduction such as principle component analysis (PCA) (Tibaduiza et al. 2013) and for supervised machine learning such as artificial neural networks (ANNs) (Hakim and Razak 2013) have been used for processing, interpretation and damage detection. In terms of model updating and system identification, a large body of literature exists pertaining specifically to structural health monitoring (SHM) of critical infrastructure. Pasquier and Smith (2015) developed a systematic system identification framework using model falsification combined with knowledge-based reasoning. They used an iterative approach in which they developed a series of physics-based candidate finite element (FE) models and compared their results with measured in-situ performance data. By applying their framework to a full-scale existing highway bridge, they confirmed its validity to perform system identification while taking into consideration systematic uncertainties. Malekzadeh et al. (2015) proposed a hybrid data analysis technique which combined model-driven and data-driven approaches. Using a small laboratory-based instrumented bridge structure, the framework consists of three sequential phases which include studying (determining structure-specific critical damage scenarios), training (running FE model simulations coupled with Monte Carlo simulations to learn relevant performance parameters), and monitoring (using trained data set to make real-time predictions). Finally, they were able to demonstrate how the developed hybrid framework might be used in future to perform real-time damage detection (facilitated via the FE model) on a full-scale bridge.

Data-centric engineering (DCE) is the emerging field of research which seeks to leverage the performance data obtained from sensors which have been installed on engineered systems and from data obtained through physics-based simulations to produce more robust predictive models. Past studies by the authors have involved developing fundamental statistical tools for analysing and processing live-streaming monitoring data and have begun to define the concept of DCE in the context of instrumented infrastructure (Lau et al. 2018a, Lau et al. 2018b).

This paper has several primary contributions, namely, it defines a new class of problems related to data-centric engineering in the context of modelling the behaviour of instrumented infrastructure. Secondly, it demonstrates a DCE

modelling approach which utilises Gaussian processes through a case study involving fibre-optic based self-sensing railway sleepers. Emerging machine learning techniques based on Gaussian process (GP) regression are nonparametric, can be used for nonlinear regression and are ideally-suited for use in structural damage detection (Fuentes et al. 2013). Multi-output Gaussian processes (Kennedy et al. 2001) have been used in a recent study concerning condition monitoring where information from two types of physics-based models (low- and high- fidelity FE models) are fused together to obtain more accurate predictions of structural responses (Zhou et al. 2016). Finally, this paper discusses the future implications and applications of DCE-based approaches for better leveraging data generated from instrumented structures.

Figure 1. Approaches for modelling engineered systems



2. Instrumented Infrastructure

DCE applications have begun to be explored within a wide-variety of sectors including cybersecurity, ecosystem modelling, extreme weather event prediction and many others. Its application to infrastructure systems, particularly those deemed as being ‘smart’ is an area of increasing interest. This paper focuses specifically on applying DCE approaches to instrumented civil engineering assets.

Case Study: Self-Sensing Railway Sleepers Installed on a Concrete Underbridge

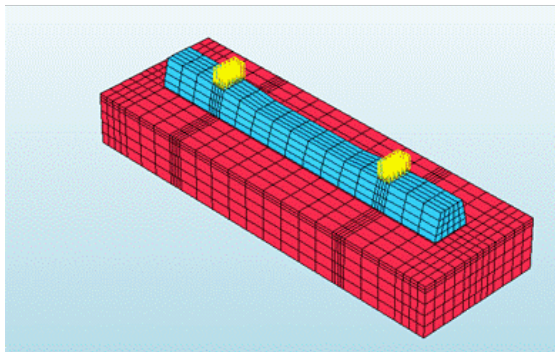
Beginning in 2015, a research project undertaken through the Centre for Smart Infrastructure and Construction at the University of Cambridge investigated the incorporation of fibre optic sensor (FOS) networks into new railway bridges and sleepers (Butler et al. 2016, Butler et al. 2017, Xu et al. 2019). This project investigated the feasibility of incorporating FOS within concrete bridge beams and sleepers during their construction and using the collected sensor data to reason about the structures’ response to passing trains. Specifically, an 11.9 metre underbridge (owned and operated by Network Rail) supported by 600 mm deep prestressed concrete beams, a 200 mm reinforced concrete deck with precast concrete parapets was constructed over a thirteen-month period (completed in

April 2016). The bridge primarily carries heavy freight trains via a single rail line over a diverted watercourse. The rail track is supported by prestressed concrete sleepers installed on 400 mm of stone ballast. Figure 2 depicts the several stages of data collected (and generated) from the self-sensing sleeper and track bed system.

Figure 2 Stages of data collection for self-sensing sleepers



a) Experimental test setup



b) Finite element model of sleeper on ballast



c) Self-sensing sleepers on bridge

To provide additional information of how the axle loads are supported by and distributed to the bridge superstructure, in addition to installing FOS within the main bridge girders and deck slab, three sleepers were instrumented and installed consecutively at the midspan of the bridge (refer to Figure 2c). In parallel to the structural monitoring programme, experimental testing was performed to simulate the loading and boundary conditions experienced by one of the installed sleepers (refer to Figure 2a). Testing was conducted at a specialized laboratory facility housed at Heriot-Watt University. A nonlinear FE model of the prestressed concrete

sleepers was constructed to provide additional information and to run parametric studies on the behaviour of the sleeper on ballast (refer to Figure 2b). Further details of the FE model and modelling process are provided in Xu et al. (2019).

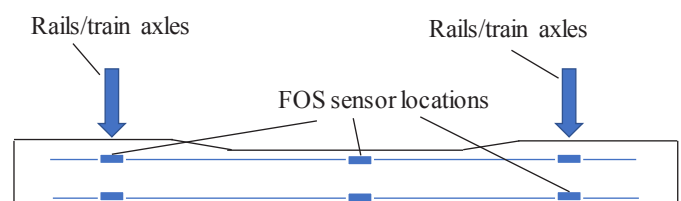
Effects of loading, ballast settlement and supporting ballast pressure were all assessed based on the experimental and numerical results. The results indicated that self-sensing sleepers in addition to being able to directly assess their own structural condition may also be used for track-level condition monitoring (e.g. ballast settlement, axle load estimation, train speed and direction indication, etc.).

FOS were installed during the sleeper manufacturing process. In total, three strain-measuring FOS (fibre Bragg grating point-based sensors) were installed along the top and bottom prestressing steel strands and embedded within the concrete (refer to To provide additional information of how the axle loads are supported by and distributed to the bridge superstructure, in addition to installing FOS within the main bridge girders and deck slab, three sleepers were instrumented and installed consecutively at the midspan of the bridge (refer to Figure 2c). In parallel to the structural monitoring programme, experimental testing was performed to simulate the loading and boundary conditions experienced by one of the installed sleepers (refer to Figure 2a). Testing was conducted at a specialized laboratory facility housed at Heriot-Watt University. A nonlinear FE model of the prestressed concrete sleepers was constructed to provide additional information and to run parametric studies on the behaviour of the sleeper on ballast (refer to Figure 2b). Further details of the FE model and modelling process are provided in Xu et al. (2019).

These sensors have a measurement accuracy of approximately $\pm 4 \mu\epsilon$ and, when used with a Micron Optics sm130 spectrum analyser can acquire data at up to 1000 Hz. All fibre optic sensor cables were routed to a central enclosure where sensor signals could be interrogated and downloaded.

All monitoring and experimental data collection activities using sensor have some associated level of uncertainty. In particular, sensor measurement accuracy may be affected by ambient temperature fluctuations, method of attachment, long-term deterioration, analyser settings and limitations, and software data interpretation algorithms. Accounting for these uncertainties in experimental and field monitoring data is critical for properly interpreting the information generated.

Figure 3 Sleeper schematic and FOS instrumentation



3. Data-Centric Engineering Approaches

This section explores how to combine the modelling elements of DCE approaches: physics-based modelling and data-driven modelling. One feature of the DCE approach is the combination of physics and statistics at the modelling stage. Another feature considers the contrast between the output of the physics-based model and the data-driven model. These two features are conceptually very different. The first fuses both models together in an effort to achieve a more representative model, whilst the second is more concerned with assessing the similarities between the separate model outputs.

We now discuss in more detail the aspects of DCE methods using the application of the railway sleeper described previously. Recall that FOS measurement data is not available along all points of the railway sleeper: data is only available at three discrete locations. In this application, the aim is to monitor the response of the *entire* sleeper, and not just at the sensor locations. The *physics* (mechanics) governing the response of the sleeper can be used to understand its response to applied axle loading between the sensor locations. One way of achieving this was explored in Gregory et al. (2018) that combines an analytical physics-based model (i.e. beams on elastic foundations) (Hetenyi 1971) and experimental data from the sleeper.

A DCE approach may combine the separate model outputs to generate predictions of the response of the sleeper to stimuli. This is in contrast to traditional modelling approaches used in engineering. In model updating for example, the predictions are obtained from a physics-based model, whilst the data is only used to tune the components of the physics-based model. Data-centric engineering advocates a more balanced approach to modelling, where each separate model can contribute to the predictions. This approach is more balanced in the following sense: first, physics-based models are typically based on a structural simplification of the asset (including loading and boundary condition assumptions). Therefore, assuming the physics-based model is the true data generation process, as in model updating, would lead to an inaccurate representation of the structures' *true* response. Conversely, the sensor data captures the response of the structure in operation using a measurement device. Thus, modelling the data directly would capture the response of the instrumented asset *and* the associated measurement errors of the sensors themselves and the measurement device (e.g. in this study, a FOS analyser). In summary, a DCE approach models the response of the structure by balancing the information generated from the physics-based model and the observed (experimental and/or operational) error-prone measurement data.

In Gregory et al. (2018), this balance is explored through multi-output Gaussian processes. A Gaussian process (Rasmussen 2004) is a model for a continuous function $f(z)$, over a problem space $z \in Z$ (e.g. space and time), that has a mean $\mu(z)$ and a covariance function $k(z, z')$, $z' \in Z$,

$$f(z) \sim \mathbf{GP}(\mu(z), k(z, z')) \quad (1)$$

A multi-output Gaussian process (Kennedy et al. 2001) is a framework for modelling more than one variable, say $f(z)$ and $u(z)$, that are connected in some way $f(z) = g(u(z))$, via Gaussian processes. The model can therefore generate predictions for any one variable, conditioned on data from all variables. In the case of the method presented in Gregory et al. (2018), each *output* corresponds to the output from the experimental testing of the sleeper or from the predictions generated from an analytic physics-based model (e.g. closed-form solution for a beam of finite length on an elastic foundation). The connection between the two sources of information in this case is the physical law connecting the two quantities measured by each of the sources: vertical deflection simulated by the physics-based model and beam curvature derived through strain measurements measured recorded during experimental testing. This approach balances the proportion of data considered from the physics-based model output with respect to the number of experimental data points. This is achieved by maximising the predictive performance of the resulting DCE model. If the physics-based model is overly simplified, this approach would prefer to use a lower proportion of data from the physics-based model. Combining the physics-based model and sensor data in this way creates a DCE-based model capable of quantifying the effects of system components that cannot be directly measured by the sensors (e.g. the stiffness of ballast supporting the sleeper).

The aforementioned study (Gregory et al. 2018) proposed one possible way of balancing the two sources of information relating to the modelled railway sleeper. However, there are many other DCE methods one could consider using for such an application. The next section outlines some future developments of these methods.

4. Future Developments of DCE Approaches

In SHM applications, there are typically three sources of information: physics-based model simulations, experimental data and data collected from the structure in operation. This raises the question of how to combine the different data sources to monitor and accurately predict the future operational response of the structure. Another issue is how to deploy results from experimental and theoretical work in operational systems (Cawley, 2018). These are important aspects of SHM where the application of DCE-based approaches is being investigated.

To illustrate the issues of combining the different sources of data used in SHM we now consider physics-based simulations in addition to experimental (see Figure 4) and operational data (see Figure 5) from the instrumented sleeper. Note that for the experimental data, the response of the sleeper is recorded under known applied forcing (i.e. simulated train axle loading). Whereas, for the operational data the response of the sleeper is recorded over time, at 50 Hz; the forcing (i.e. the actual axle loads) is unknown.

The physics-based model can generate simulated responses for a variety of variables, such as time and forcing. The manner in which the experimental and operational data are combined will depend on the input of the selected physics-based model, i.e. response as a function of space and force, or space and time. Figure 6 displays the information from the three sources in response (beam curvature) and space.

Note that the physics-based model is based on several simplifications. The form of the analytic physics-based model used in Gregory et al. (2018), assumes a constant flexural rigidity of the railway sleeper. This assumption is unrealistic as can be seen in Figure 6 from the anti-symmetry in the experimental and operational data at the sensor locations, in contrast to the physics-based model simulations. The Figure displays the empirical variation in the experimental and operational data, which does not exist in the physics-based simulations since they are deterministic.

Figure 4 Experimentally-measured curvature for the three sensors (black & blue = sensors under loading points; red = midspan sensors) (curvature vs. time).

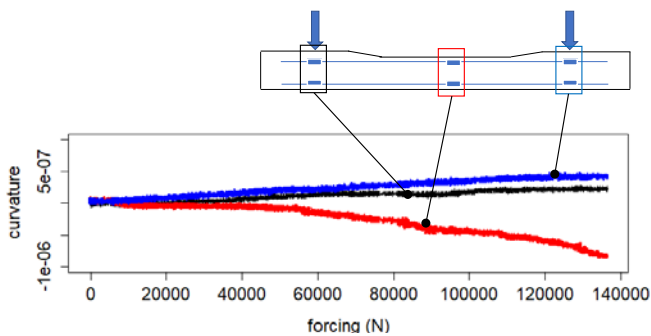
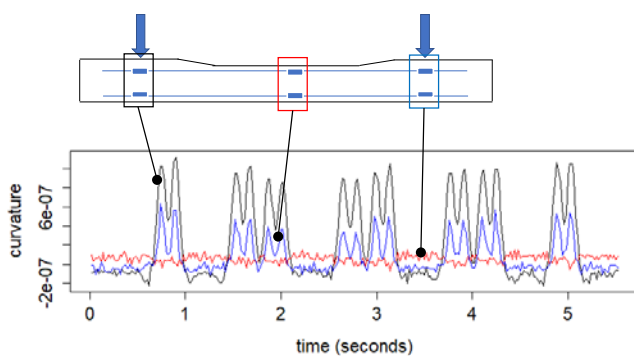


Figure 5 Measured operational curvature values for the three sensors (represented by different colours) on a sleeper from Bridge 11 during a train passage event – (curvature vs. sleeper length).



Recall from the start of the section, that the aim of using DCE in this study is to predict the operational response of the structure, by leveraging the data collected from experiments and physics-based models. In addition to this, it is of interest to quantify the differences between the experimental and operational responses of the sleeper. We now give an example

of how the DCE model, discussed in the previous section, can be used for this purpose.

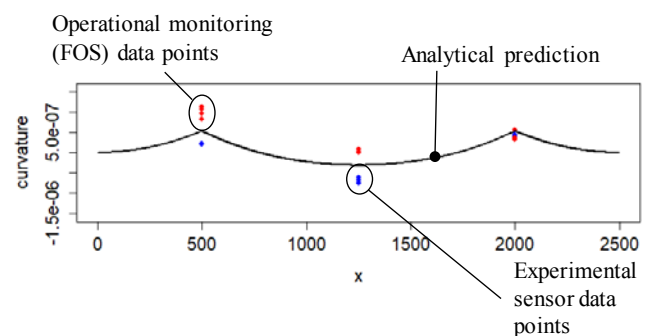
A way of combining all three sources of information for the railway sleeper application using the DCE model proposed in Gregory et al. (2018) is to employ a connecting equation between the experimental and operational data. An example of this connecting equation between the two sources of data is,

$$\text{operational data} = (\rho \times \text{experimental data}) + \xi, \quad (2)$$

Where ρ is an unknown scaling parameter which needs to be estimated and ξ is a random noise term (with potentially unknown parameters that need to be estimated). Connecting all three sources of information can then be done by utilising the aforementioned connection between the experimental data and the physics-based model in addition to the one in (2). The output (e.g. predictions of operational structural responses) produced from extending the DCE model in this way, will now be based on all three data sources (i.e. field monitoring, experimental, and the analytical model) (illustrated in Figure 6). In principle, this extension to the Gaussian process framework can be used to leverage both experimental data and physics-based model simulations. However, the performance of the resulting DCE method depends on the choice of the connecting equation.

Instrumented infrastructure, such as the railway sleeper considered throughout this paper, produce complex data sets, i.e. high dimensional and high frequency. Many traditional engineering approaches, including FE modelling, are inadequate to deal with such data in real-time. For example, it may be necessary to monitor some structures continuously. This presents a computational challenge when the data is revealed at a high frequency – it would be infeasible to recalibrate or run a FE model at 50 Hz. Data-centric engineering methods seek to address this difficult and unresolved problem.

Figure 6 Physical simulation of curvature (black), taken with forcing of standard rail carriage (125,000N), operational curvature data (red), taken at the peak of curvature from a carriage going over it (e.g. 0.75 seconds from Figure 5), and experimental data (blue), taken at a forcing of 125,000 N.



5. Conclusions

This paper has provided an overview of how the emerging analytical approach of DCE can be applied to instrumented infrastructure systems. By gathering quantifiable performance data from critical bridges, buildings, tunnels etc. and combining this information with traditional numerical (i.e. FE) modelling approaches, DCE seeks to create more robust and balanced predictive models. A contrast was drawn between data-driven models, model updating (and system identification) and data-centric models. A DCE model proposes a more balanced approach to modelling, where each separate model (i.e. data-driven and physics-based) can contribute to the predictions. Thus, by discussing the subtleties of this distinction, this study has provided a more comprehensive definition of DCE. An example of a DCE approach using a two-stage Gaussian process, was applied to fuse experimental and field monitoring sensor data captured from instrumented railway sleepers. Several key advantages of a DCE-based approach over traditional data-driven, model-based and system identification approaches include:

- The development of a separate predictive model that is not based on data-driven and physics-based (i.e. FE) models alone which is capable of accounting for uncertainties in model assumptions and measurement error,
- Allowing for the systematic balancing of inputs from the data-driven and physics-based model results; and
- The ability to be readily adapted to combine (and balance) inputs from multiple sources of data (e.g. from experimental tests, field monitoring and numerical simulations).

There are still many complex data science and engineering challenges to address including:

- Data curation and processing: all data from an instrumented infrastructure cannot not be stored given the high-frequency recording and the dimensionality of the data. This raises questions of the sort: Which parts of the data should be stored? How is the “importance” of the kept parts measured?
- False positives over infinite/long time horizons: in the setting of a structure instrumented with a sensor network, it is feasible that data will be accrued continuously for very long periods. Any system/methodology that is used to signal changes/degradation/damage will produce false signals. A challenge, from a statistical point-of-view, is controlling the false signal rates over infinite time horizon.
- Integrating live-streaming statistical models (based on continuously recorded sensor data) with physics-based models (e.g. finite element); and
- Demonstrating the performance of DCE models as applied to full-scale and operational engineering systems.

The flexibility of DCE approaches for combining a variety of different data sets (and models) provides a promising framework for tackling these challenges.

Acknowledgements

The authors gratefully acknowledge the EPSRC and Innovate UK, the Cambridge Centre for Smart Infrastructure and Construction (CSIC) Innovation and Knowledge Centre (EPSRC grant reference number EP/L010917/1) and the Lloyd’s Register Foundation Programme on Data-Centric Engineering at the Alan Turing Institute for supporting this project; the invaluable use and facilitation of Heriot-Watt University testing facilities; the assistance with self-sensing sleeper fabrication from CEMEX; the on-site instrumentation assistance of J. Shardelow and N. Gibbons of the University of Cambridge; and the experimental testing and finite element modelling expertise offered by J. Xu of the Harbin Institute of Technology.

References

- American Society of Civil Engineers. 2017 Infrastructure Report Card. Washington D.C., 2017.
- Bowers et al. (2016). Smart Infrastructure – Getting More from Strategic Assets. 7 pgs. Centre for Smart Infrastructure and Construction, Cambridge UK.
- Butler LJ, Gibbons N, Ping H, Elshafie MZEB, and Middleton CR (2016). Evaluating the early-age behaviour of full-scale prestressed concrete beams using distributed and discrete fibre optic sensors. *Journal of Construction and Building Materials*, 126: 894 – 912.
- Butler LJ, Xu J, Ping H, Gibbons N, Dirar S, Middleton CR, and Elshafie MZEB (2017). Robust fibre-optic sensor arrays for monitoring early-age performance of mass-produced concrete railway sleepers. *Journal of Structural Health Monitoring* 17(3): 635 – 653.
- Canadian Construction Association. The Canadian Infrastructure Report Card 2016. Canadian Society for Civil Engineering & Federation of Canadian Municipalities. 2016.
- Cawley P (2018). Structural health monitoring: Closing the gap between research and industrial deployment, *Structural Health Monitoring*, 17(5): 1225 – 1244.
- Farrar CR and Worden K (2013). *Structural health monitoring: a machine learning perspective*. John Wiley & Sons, Chichester, U.K. 643 pgs.
- Farrar CR and Worden K (2013). *Structural health monitoring: a machine learning perspective*. John Wiley & Sons, Chichester, U.K. 643 pgs.
- Fuentes R, Cross EJ, Halfpenny A, Barthorpe RJ and Worden K (2014). Autoregressive Gaussian processes for structural damage detection. *Proceeding of the international conference on uncertainty in structural dynamics*, Leuven, Belgium: 469 – 483.
- Gregory A, Lau D-H, Girolami M, Butler L, Elshafie M. (2018) The synthesis of data from instrumented structures and physics-based models via Gaussian processes. arXiv:1811.108

Hakim SJS and Razak HA (2013). Structural damage detection of steel bridge girder using artificial neural networks and FE models. *Steel and composite structures* 14(4):367 – 377.

Hetenyi M (1971) *Beams on elastic foundation: theory with applications in the fields of civil and mechanical engineering*. University of Michigan.

Lau FD-H, Butler LJ, Adams NM, Girolami MA, Elshafie MZEB. (2018a). The role of statistics in data-centric engineering. *Statistics and Probability Letters, Special Issue on the Role of Statistics in the Era of Big Data*. 136.

Lau FD-H, Butler LJ, Adams NM, Elshafie MZEB, Girolami MA. (2018b). Real-time statistical modelling of data generated from self-sensing bridges. *ICE Journal of Smart Infrastructure and Construction*. 171(1): 3 -13.

Malekzadeh M., Atia G., & Catbas F.N. (2015). Performance-based structural health monitoring through an innovative hybrid data interpretation framework. *J Civil Struct Health Monit* 5:287 – 305.

Pasquier R. and Smith I.F.C. (2016). Iterative structural identification framework for evaluation of existing structures. *Engineering structures* 106:179 – 194.

Rasmussen, CE (2004). *Gaussian processes in machine learning. Advanced lectures on machine learning*. Springer, Berlin, Heidelberg: 63-71.

Tibaduiza DA, Mujica LE and Rodellar J (2013). Damage classification in structural health monitoring using principle component analysis and self-organizing maps. *Structural control and health monitoring* 20: 1303 – 1316.

Xu J, Butler LJ, Elshafie MZEB. (2019). Experimental and numerical investigation of the performance of self-sensing concrete sleepers. *Structural health monitoring*.

Zhou K, Liang G, and Tang J (2016). Vibration analysis of structure with uncertainty using two-level Gaussian processes and Bayesian inference. *Journal of Physics: Conference Series*. Vol. 744.